

GlyGen

Glycoscience data integration: Data acquisition and standardization

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

PI: Raja Mazumder
The George Washington University



PI: Michael Tiemeyer
CCRC, University of Georgia

NIH Common Fund Glycoscience Program Meeting | June 16-17 2020



Web portal: <https://glygen.org>
Data store: <https://data.glygen.org>
WS API: <https://api.glygen.org>

CONTACT
Raja Mazumder mazumder@gwu.edu
Michael Tiemeyer mtiemeyer@ccrc.uga.edu

GlyGen.org

https://twitter.com/gly_gen



Conflict of interest

Nothing to declare



Harvard University
Richard Cummings
Akul Mehta

The Jackson Laboratory
Judith Blake

Soka University
Kiyoko Aoki-Kinoshita

The Griffith University
Matthew Campbell
Jodie Abrahams

University of Delaware
Vijay Shankar
Peng Su

Imperial College London
Ten Feizi
Yukie Akune

Macquarie University
Nicki Packer

University of Georgia

Michael Tiemeyer
René Ranzinger
William York
Michael Pierce
Robert Woods
Rupali Mahadik
Tatiana Williamson
Sena Arpinar
Sanath Bhatt
Sujeet Kulkarni
Sandeep Nakarakkumula
Gaurav Agarwal
Vinamra Jain

EMBL-EBI

Maria Martin
Preethi Vasudev
Leyla Jael Garcia Castro

NIH-NCBI-NCI

Kim Pruitt
Evan Bolton
Jefferey Gildersleeve
Paul Thiessen
Terrence Murphy

The George Washington University

Raja Mazumder
Robel Kahsay
Jeet Vora
Rahi Navelkar
Nagarajan
Pattabiraman
Gareth Owen
Xiying Ding
Reza Mousavi
Xavier Holmes
Brian Fochtman
Tyler Stewart

Georgetown University

Nathan Edwards
Radoslav Goldman
Darren Natale
Karen Ross
Wenjin Zhang

Introduction to GlyGen

Progress in Data Acquisition and Data Accessibility

Progress in Connecting with Other Resources

Progress in Collaborations with Subcontracts

Progress in Portal Enhancement

Progress in Outreach

Year 4 Goals

GlyGen

Introduction

Computational and Informatics Resources for Glycoscience

GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.

GlyGen is an international project funded by The National Institutes of Health

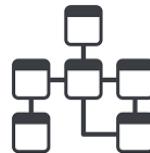
York, W. S., Mazumder, R., Ranzinger, R., Edwards, N., Kahsay, R., Aoki-Kinoshita, K. F., ... Zhang, W. (2019). GlyGen: Computational and Informatics Resources for Glycoscience. *Glycobiology*. doi: 10.1093/glycob/cwz080

Aim 1 – Integrate data and knowledge from diverse disciplines relevant to glycobiology.

Aim 2 – Implement a web portal for browsing, searching, and visualization of knowledge relevant to glycobiology.

Aim 3 – Develop and integrate critical new glycoinformatics infrastructure.

Aim 4 – Engage the community in software planning and development.



Data collection

User-friendly
Interface

Easy Data access

Streamlined
Data Integration

Uses-cases
oriented
searches

Documentation

Quality control
and assurance

Queries across
multiple
domains

Community
research

User research



Use cases gathered

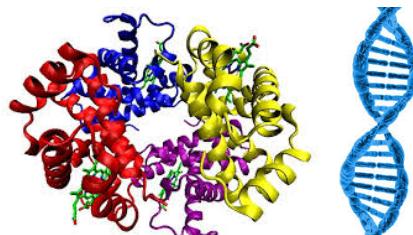
As a (*user profile*)
I want (*requirement*)
So that (*benefit*)

Prioritization & Validation

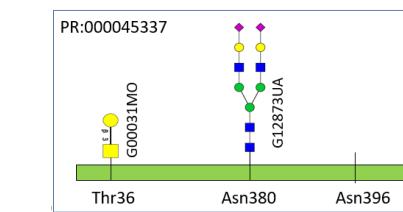


Three Data Domains in GlyGen

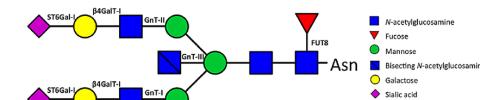
Protein
Centric



Protein-Site
Centric



Glycan
Centric

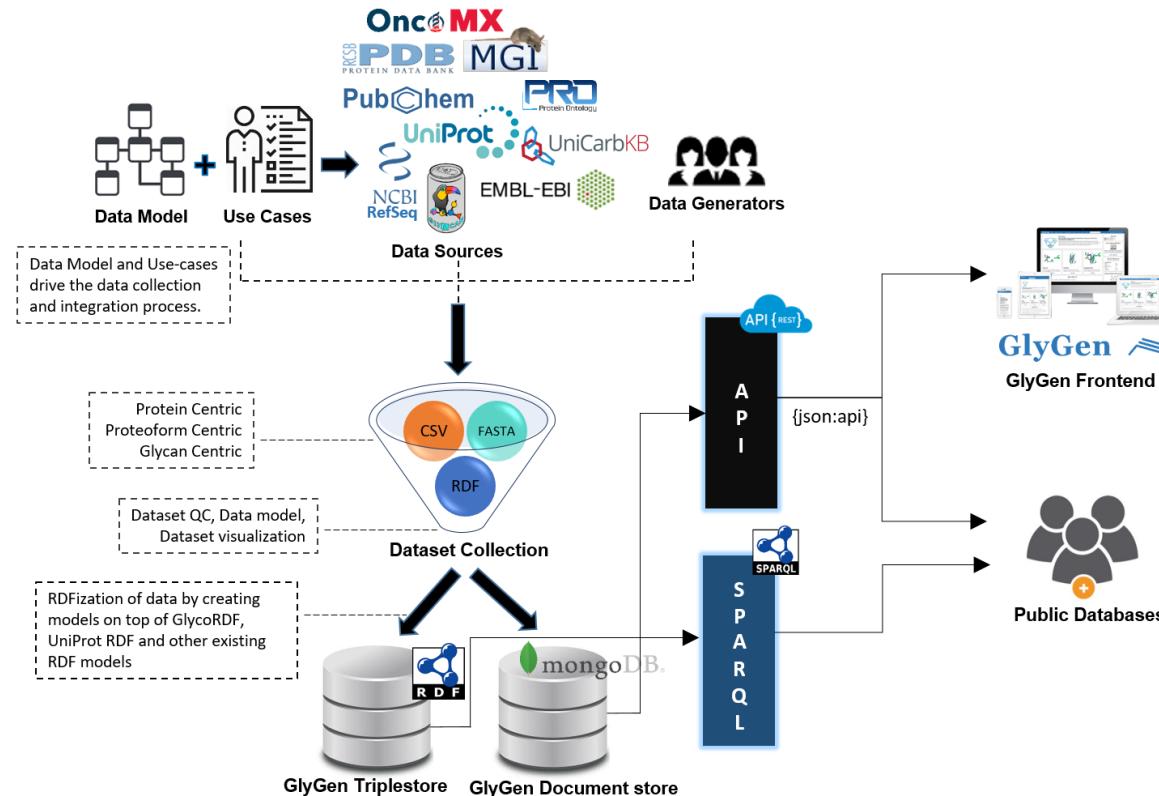


Protein centric - Data types and information that is about some particular protein coding gene, or that can be mapped to the canonical protein sequence representing that gene.

Protein-Site centric - Data types and information of proteins that are glycosylated at specific sites with or without glycan information.

Glycan centric - Data types and information that is about specific glycans.

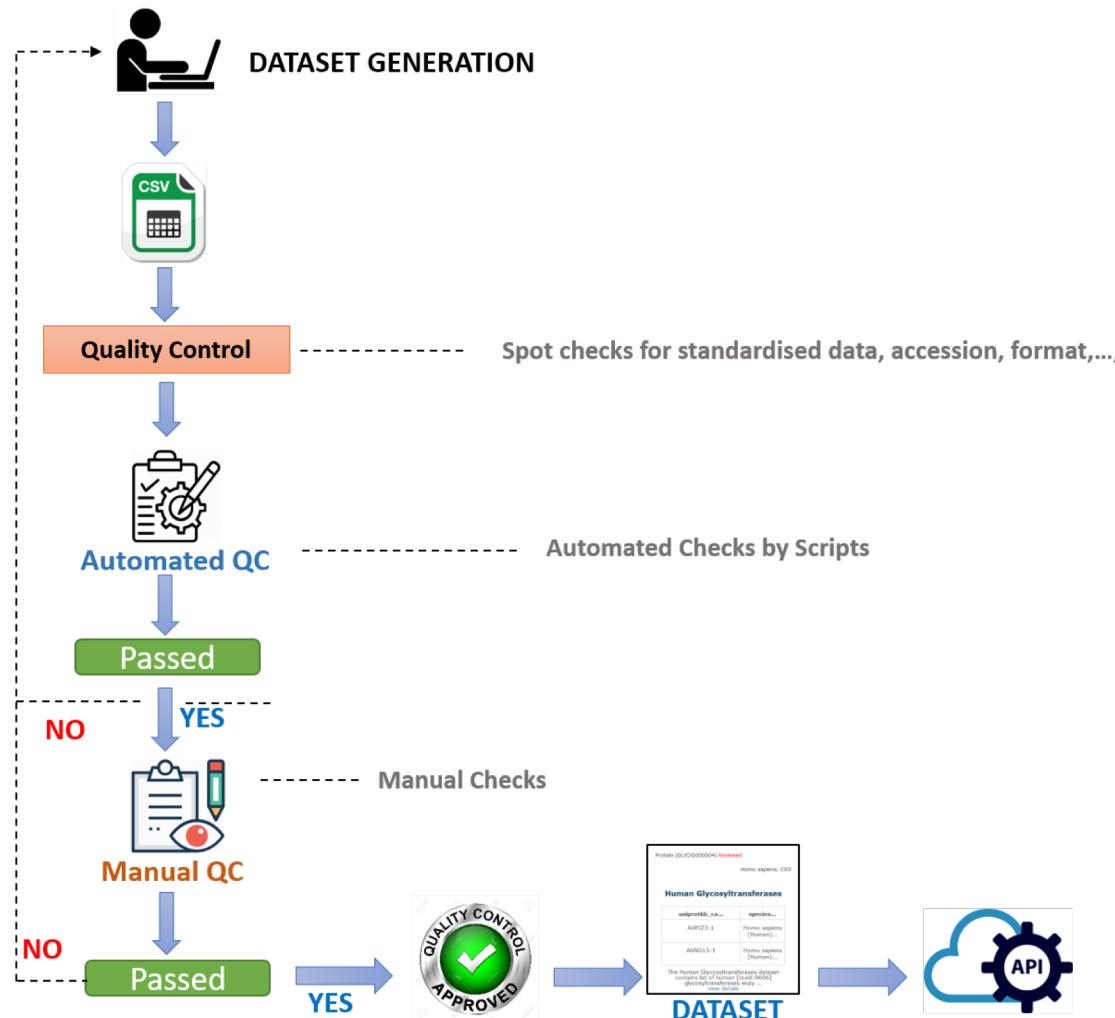
GlyGen Data Collection and Integration

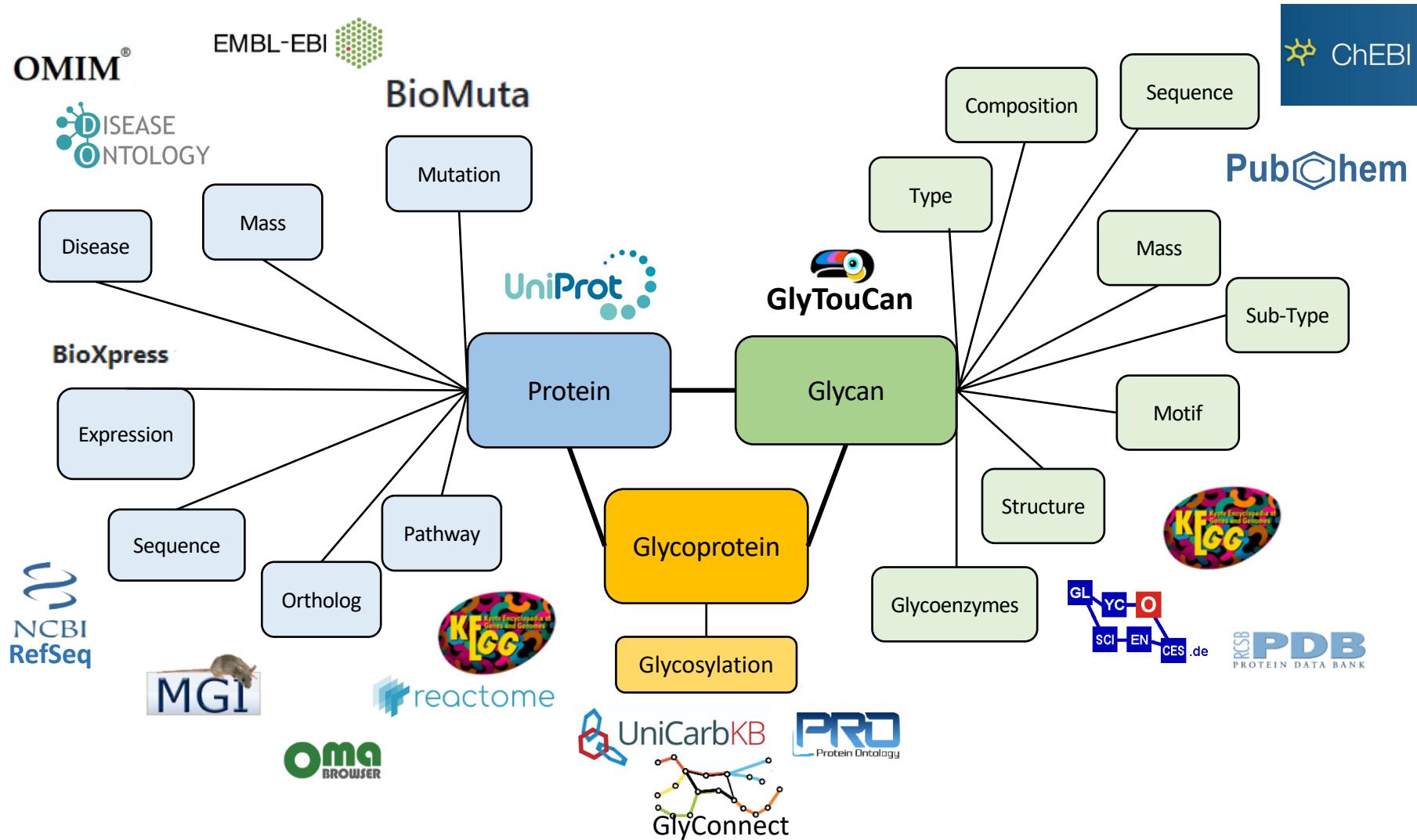


GlyGen's
Data Sources



GlyGen.org





GlyGen Portal – List and Detail Pages

Protein Search / Protein List

Summary of your Protein Search

Performed on: February 21st 2020, 10:41:08 am (EST)

Organism: Homo sapiens

[Update Results](#) [Modify Search](#)

** To perform the same search again using the current version of the database, click "[Update Results](#)".

Page [Prev](#) [1](#) [Next](#) Records per page [20](#)

"20656 Proteins were found"

[DOWNLOAD](#)

GlyGen Data Portal – <https://data.glygen.org>

Dataset Collection Website 1.5.7 | Data 1.5.13

Search term [Search](#)

You can search by BCO ID, name or contributor.

 **BioCompute Objects** [Home](#) | [Integration Workflow](#) | [FAQ](#) | [Release History](#)

GLYDS000001 sample view

Human Proteome Masterlist (UniProtKB)
 The Human Proteome Masterlist dataset contains human [taxid:9606] UniProtKB canonical accessions mapped to the reviewed (SwissProt) and unreviewed (TrEMBL) UniProtKB isoforms via gene grouping. The dataset is derived from 2019_09 UniProtKB release. If you use this dataset please provide proper attribution to EMBL-EBI-UniProtKB and GlyGen

Version: v-1.5.13 01/14/2020 [BCO JSON](#) [DOWNLOAD](#)

uniprotkb_canonical_ac	status	gene_name	reviewed_isoforms	unreviewed_isoforms
Q9Y6C9-1	reviewed	MTCH2	Q9Y6C9-1	E9PIE4-1
Q658T7-1	reviewed	FAM90A2P	Q658T7-1	
Q658T7-1	reviewed	FAM90A2P	Q658T7-2	
P30613-1	reviewed	PKLR	P30613-1	A0A0G2JLC7-1
P30613-1	reviewed	PKLR	P30613-2	F8W6W2-1

APIs and RDF SPARQL endpoint



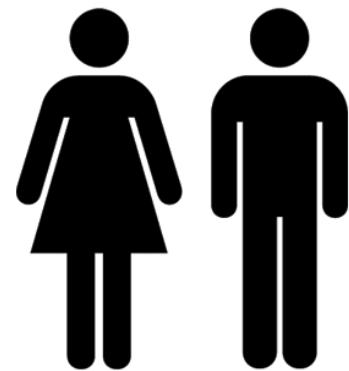
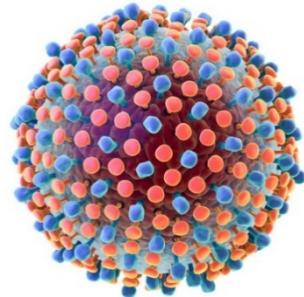
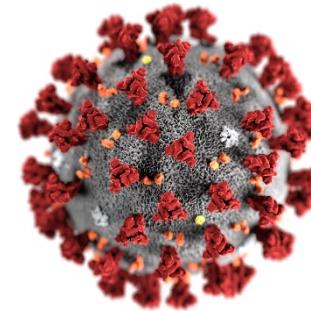
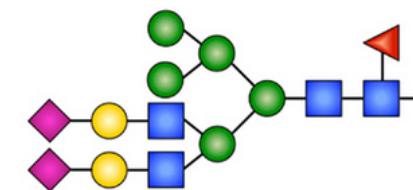
<https://sparql.glygen.org>

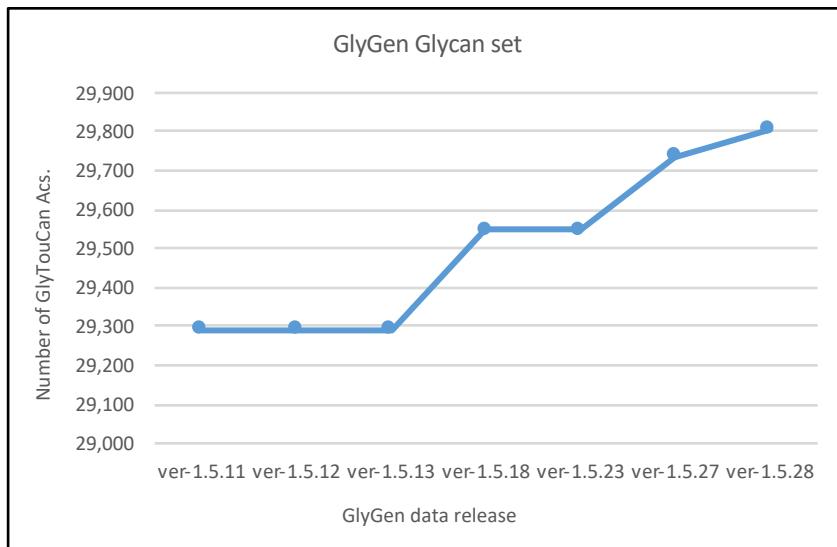
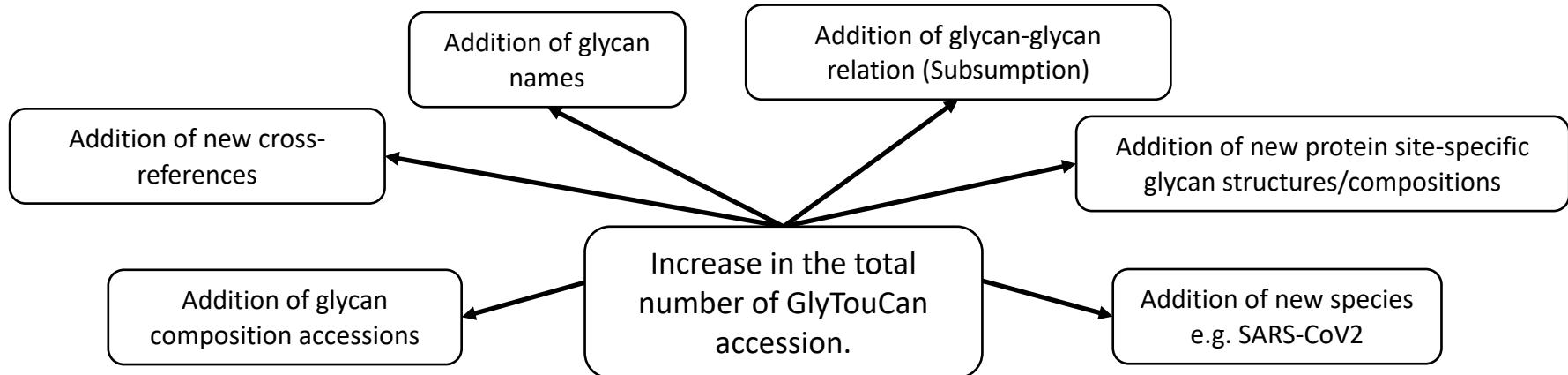


<https://api.glygen.org>

GlyGen

**Progress in Data Acquisition
and Data Accessibility**

Year 1 - Human**Year 1 - Mouse****Year 2 - Rat****Year 3 -
HCV 1a & HCV 1b****Year 3 -
SARS-CoV1 & SARS Cov2****Year 3 –
Synthetic Glycans**



<https://data.glygen.org/upload>

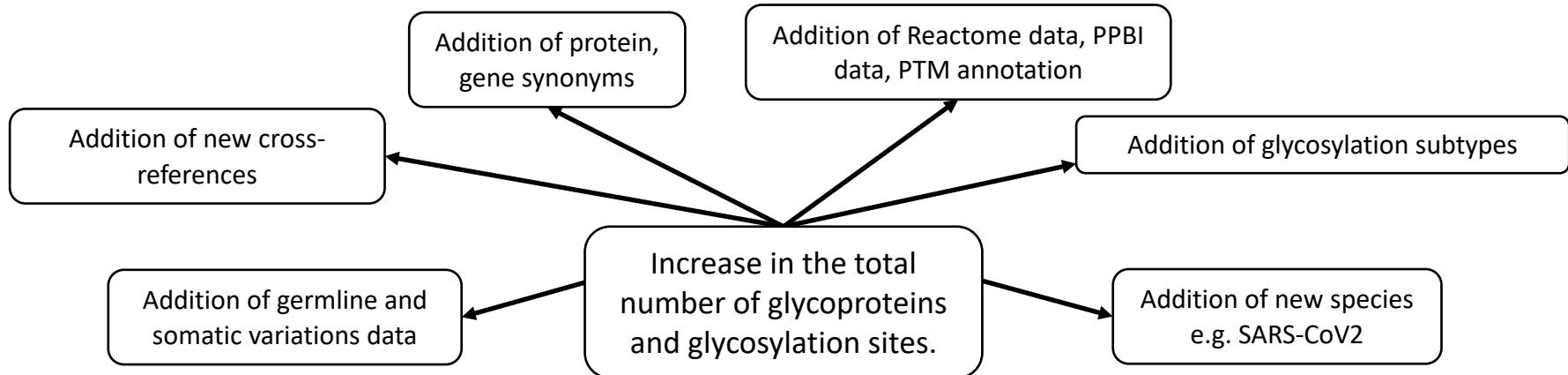
Session A. Use case and data submission

The screenshot shows the 'Dataset Collection' section of the GlyGen portal. It features a search bar labeled 'Search term' and a 'Search' button. Below the search bar, there is a note: 'You can search by BCO ID, name or contributor.' At the bottom of the page, there are links for 'Home', 'Integration Workflow', 'FAQ', 'Release History', and 'Upload File'.

GlyGen



Protein Data Expansion in Year 3



Species	Glycoproteins	Glycosites
<i>Homo sapiens</i>	5,248	21,256
<i>Mus musculus</i>	3,884	14,386
<i>Rattus norvegicus</i>	2,108	7,493
SARS coronavirus SARS-CoV-2 or 2019-nCoV)	2	26
SARS coronavirus (SARS-CoV-1)	2	52
Hepatitis C virus (genotype 1a, isolate H)	1	15
Hepatitis C virus (genotype 1b, isolate Japanese)	1	20



O-GlcNAc protein data

GAG Interaction data

Variation data for viruses

O-Glycoproteome data

Lectin data

Glycogenes data

Glycan recognition data



New disease data

Update to existing data

Mutagenesis data

PRO data

New use case data

Literature mining data

New Glycosylation data

GlyGen

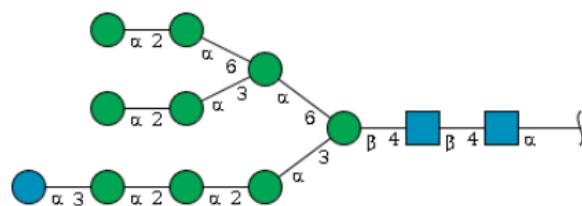


Progress in Connecting with other Resources

Using PubChem CID as an anchor database to register/map GlyTouCan glycan entries in ChEBI.

CID 91850660

PubChem



Extracting the existing x-ref to ChEBI or downloading structure data from PubChem to generate SDF format files for ChEBI submissions.

GlyTouCan

G24620AI

ChEBI

CHEBI:145597

Currently, over 6000 GlyTouCan accessions are mapped to ChEBI IDs.
Over 1700 added in PubChem through ChEBI monthly submissions.

See: <https://data.glygen.org/GLYDS000296>

GlyGen URL	https://glygen.org/glycan/G24620AI
GlyGen Classification	N-glycan
GlyGen Motifs	N-Glycan core basic (is reducing end) N-Glycan high mannose (is reducing end)

▼ GlyGen

Source: GlyGen
Record Name: G24620AI
URL: <https://glygen.org/glycan/G24620AI>
Description: GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. The web portal allows users to perform unique searches that cannot be executed in any of the integrated databases alone.
License Note: CC BY 4.0
License URL: <https://www.glygen.org/license.html>
<https://pubchem.ncbi.nlm.nih.gov/compound/91850660>

PubChem



GlyGen

General

- GlyToCan Accession: [G24620AI](#)
- Monoisotopic Mass: 2,044.7 Da (Permethylated: 2577.29 Da)
- Composition: Hex₁₀ HexNAc₂
- Glycan Type/Subtype: N-glycan
- InChI Key: [OYKAKPUDUALZCM-QESMDMLYSA-N](#)

<https://glygen.org/glycan/G24620AI>

LinkOut to external resources

Dryad Digital Repository

[Dryad Digital Repository]

Ensembl

[Ensembl]

GlyGen

[GlyGen]

epidermal growth factor receptor protein

[Guide to Pharmacology]

https://www.ncbi.nlm.nih.gov/protein/NP_005219.2

NCBI RefSeq/Protein



General

- Gene Name: ERBB; ERBB1; HER1; mENA; NISBD2; PIG61; ERBB; HER1; ERBB1; EGFR
- Gene Location: Chromosome: 7 (55,019,021 - 55,211,628)

Ensembl Gene [1](#)

- UniProtKB ID: EGFR_HUMAN
- UniProtKB Accession: [P00533-1](#)
- UniProtKB Accession Length: 1210
- UniProtKB Entry Name: Epidermal growth factor receptor; 2.7.10.1; Receptor tyrosine-protein kinase erbB-1; Proto-oncogene c-ErbB-1
- Chemical Mass: 134,277 Da
- RefSeq Accession: [NP_005219.2](#)
- RefSeq Name: epidermal growth factor receptor isoform a precursor
- RefSeq Summary: The protein encoded by this gene is a transmembrane glycoprotein that is of the protein kinase superfamily. This protein is a receptor for members of the epidermal growth factor family. EGFR is a cell surface protein that binds to epidermal growth factor. Binding of the protein ligand induces receptor dimerization and tyrosine autophosphorylation and leads to cell proliferation. Mutations in this gene are associated with lung cancer. [provided by RefSeq, Jun 2016].

<https://glygen.org/protein/P00533>



Submitting annotations to specific UniProtKB entry pages



Integration in GlyGen



Glycosylation ⁱ	305	N-linked (GlcNAc...) asparagine; by host	Sequence analysis
Glycosylation ⁱ	385	O-linked (Hex...) threonine; by host	1 Publication
Glycosylation ⁱ	396	O-linked (Hex...) threonine; by host	1 Publication
Glycosylation ⁱ	401	O-linked (Hex...) serine; by host	1 Publication
Glycosylation ⁱ	404	O-linked (Hex...) serine; by host	1 Publication
Glycosylation ⁱ	417	N-linked (GlcNAc...) asparagine;	
Glycosylation ⁱ	423	N-linked (GlcNAc...) asparagine;	
Glycosylation ⁱ	430	N-linked (GlcNAc...) asparagine;	
Glycosylation ⁱ	448	N-linked (GlcNAc...) asparagine;	
Glycosylation ⁱ	473	O-linked (Hex...) threonine; by h	
Glycosylation ⁱ	518	O-linked (Hex...) threonine; by h	

Manual assertion based on experiment inⁱ

"Mass spectrometric analysis of hepatitis C viral envelope protein E2 reveals extended microheterogeneity of mucin-type O-linked glycosylation."
Braeutigam J., Scheidig A.J., Egge-Jacobsen W.
Glycobiology 23:453-474(2013) [PubMed] [Europe PMC]
[Abstract]
Cited for: GLYCOSYLATION AT THR-385; THR-396; SER-401; SER-404; THR-473 AND THR-518; IDENTIFICATION BY

Submitted evidence annotation for 2 proteins entries and 16 glycosylation sites. More to follow



GlyGen's QA process for ensuring that human, mouse and rat-specific glycans are validated has conversely helped GlyTouCan in turn

GlyGen's GNOME annotation project has similarly helped GlyTouCan reconsider how subsumption can be handled

GlyGen's motif list will also be incorporated in the future, providing a curated and updated list of biologically important motifs

GlyGen

Progress in Collaborations with Subcontracts

- New protein datasets for
 - Rattus Norvegicus
 - HCV 1a
 - HCV 1b
 - SARS-CoV-1
 - **SARS-CoV-2**
- Updated SARS-CoV-2 dataset as and when UniProt pre-release is updated
- dbSNP / variant TSV datasets for human, mouse & rat
- Rhea reaction info - id, equation & description
- Binary interaction data - more info like subunit detail
- Attribution (source & evidence) for all annotations
- Enhanced stats and quality checks
- Contribution for GlyGen Data Paper - Supplementary document
- Providing documentation for the dataset and GlyGen FAQs
- Datasets generation hand-in-hand with UniProt release (11 in 2019, 2 in 2020) (ongoing)
- GlyGen x-ref from UniProt (in progress)



Maria Martin
Preethi Vasudev

P00533-1 EGFR Loss of Glycosylation @ Position 363 (NCT>NCI)

General

Species

Function

GO Annotation

Glycosylation

Sequence

Pathway

Isoforms

Sequence

ProtVista

N-linked Sites 16

O-linked Sites

Mutation 45

Sequon 15

	+10	+20	+30	+40	+50
1	M RPSGTAGAA	LLALLAALCP	ASRALEEKV	CQGTSNKLTQ	LGTfedHFLS
61	V LGNLEITYV	QRNYDLSFLK	TIQEVAGYVL	IAALNTVERIP	LENLQIIRGN
121	VLSNYDANKT	GLKELPMRNL	QEILHGAVRF	SNNPALCNVE	SIQWRDIVSS
181	QNHLGSCQKC	D PSCPNGSCW	GAGEENCQKL	TKIICAQQCS	GRCRGKSPSD
241	TGPRESDCLV	CRKFREATC	K DTCPPMLY	NPTTYQMDVN	PEGKYSFGAT
301	VTDHGSCVRA	CGADSYEMEE	DGVRKCKCE	GPCRKVCNGI	GIGEFKD SLS
361	NCTS	SGDLH	FTHTPPLDPQ	ELDIILKTVK	ITGFLLIQAW
421	ENLEIIRGRT	KQHGQFSLAV	VSLNITSGL	RSLKEISDGD	VIISGNKNLC
					YANTINWKKL

General

Species

Function

GO Annotation

Glycosylation

Sequence

Pathway

Isoforms

Homologs

Disease

Mutation

Expression Tissue

Expression Disease

Mutation

Sources	Annotation name	Disease	Start pos	End pos	Sequence	Type
BioMutant 1	Somatic mutation passed 1 out of 6 filters: n-glyco-site-loss (NCT->NCI).	Brain cancer (DOID: 1319)	363	363	T → I	Point mutation
dbSNP 1	Germline mutation passed 1 out of 1 filters: n-glyco-site-loss (NCT->NCI).	None (DOID: None)	363	363	T → I	Point mutation
dbSNP 1	Germline mutation passed 1 out of 1 filters: n-glyco-site-loss (NRT->SRT).	None (DOID: None)	413	413	N → S	Point mutation
dbSNP 1	Germline mutation passed 1 out of 1 filters: n-glyco-site-loss (NRT->NRK).	None (DOID: None)	415	415	T → K	Point mutation

P21802-1 FGFR2 Gain of Glycosylation @ Position 654 (NNI > NNT)

Sequence

ProtVista

+10 +20 +30 +40 +50

```

1 MVSWGRFICL VVVTMATELSL ARPSFSLVED TTLEPEEPPT KYQISQPEVY VAAPGESLEV
61 RCLLKDAAVI SWTKDGVHLG PNNRTVLIGE YLQIKGATPR DSGLYACTAS RTVDSETWYF
121 MVNVTDAISS GDDDEDDTDGA EDFVSESENSNN KRAPYWTNTTE KMEKRLHAVP AANTVKFRCP
181 AGGNPMPTMR WLKNGKEFKQ EHRRIGGYKVR NQHWSLIMES VVPSDKGNYT CVVENEYGSI
241 NHTYHLDVVE RSPHRPILQA GLPANASTVV GGDVEFVCKV YSDAQPHIQW IKHVEKNGSK
301 YGPGLPYLK VLKAAGVNNTT DKEIEVLYIR NVTFEDAGEY TCLAGNSIGI SFHSAWLTVL
361 PAPGREKEIT ASPDYLEIAI YCIGVFLIAC MVTIVILCRM KNTTKKPDFS SQPAVHKLTK
421 RIPLRRQVTV SAESSSSMNS NTPLVRITTR LSSTADTPML AGVSEYELPE DPKWEFPRDK
481 LTLGKPLGEG CFGQVVMAEA VGIDKDKPKE AVTVAVKMLK DDATEKDLSL LVSEMEMMMK
541 IGKHKNIIINL LGACTQDGPL YVIVEYASKG NLREYLRARR PPGMEYSYDI NRVPEEQMTF
601 KDLVSCTYQL ARGMEYLASQ KCIHRDLAAR NVLVTENNVM KIADFGLAR INNIIDYYKT
661 TNGLPVKWM APEALFDRVY THQSDVWSFG VLMWEIFTLG GSPYPGPVPE ELFKLKKEGH

```

 N-linked Sites 8 O-linked Sites 1 Mutation 32 Sequon 11

Mutation

Sources	Annotation name	Disease	Start pos	End pos	Sequence	Type
dbSNP 1	Germline mutation passed 1 out of 1 filters: n-glyco-sequon-gain (NNI->NNT).	None (DOID: None)	654	654	I → T	Point mutation

- ~ 30K GlyTouCan accessions
- Interactive web-site, with search capability
 - Plus triple-store with curated queries.
 - Versioned, exported tables (w/ release tag)
- Semantic annotations:
 - Glycan classification, Glycan motif alignments,
 - Subsumption based species inference
 - Glycan names, Composition strings
 - GNOme and GlycoMotif integration



Nathan Edwards
Wenjin Zhang

<http://grg.tn/GlycanData>

- GlycoMotif aggregates and organizes more than 659 motifs from many collections
 - Accessible interactively and via triple-store
 - NEW! Alignments of all motifs to GlyTouCan glycans
 - Curated GlyGen Motifs
- Basis for GlyGen glycan classifications

<http://grg.tn/GlycoMotif>

Collections

Name	Motifs
CCRC Motifs	116
GlyGen Motifs	141
GlyTouCan Motifs	62
GlycoEpitope Epitopes	172
Glydin	533
Glydin - BiOligo	304
Glydin - Cermav	126
Glydin - Cummings	121
Glydin - Hayes	165
Glydin - SugarBind	196
UniCarbKB Motifs	146
All Motifs	659

Glycan Naming and Subsumption Ontology

- **GNOme** organizes GlyTouCan glycan accessions:
 - for **interactive browsing**,
 - for **text-based searching**, and
 - for **automated reasoning**
 - with well-defined characterization levels.
- All of GlyTouCan, but restrictions for GlyGen and BCSDB (bacteria) also provided.
- Recommended glycan identifier for glycopeptides in mzIdentML.

<http://gnome.glyomics.org>

GlyTouCan/GlyCosmos (Soka University)

- GlyGen has been added as a GlyTouCan partner so that links to GlyGen from any GlyTouCan entry page can be added automatically
- GlyTouCan subsumption data is being tested with the help of GlyGen's feedback
- GlyGen's GNOME browser has been added to GlyCosmos
- Glycoprotein data from GlyCosmos have been incorporated into GlyGen
- Links to GlyGen are being added to GlyCosmos Glycoprotein entry pages
- GlyCosmos SPARQLList server has been especially set up for GlyGen to easily allow better SPARQL access to GlyCosmos triplestore data
- Other data, such as glycome information from GlycomeAtlas, have been provided from GlyCosmos, with more on the way



Kiyoko Aoki-Kinoshita



Protein Ontology (Georgetown University)

GlyGen

Adapted PRO framework to represent glycosylated proteoforms, including representation of complex glycosylated forms (multiple modification sites, ambiguity)

Improved data exchange protocol to ensure that UniCarbKB data used as input into PRO accurately reflects the information in UniCarbKB

Provided advice on development of the Glycan Naming and Subsumption Ontology (GNOme), and facilitated its inclusion into the OBO Foundry

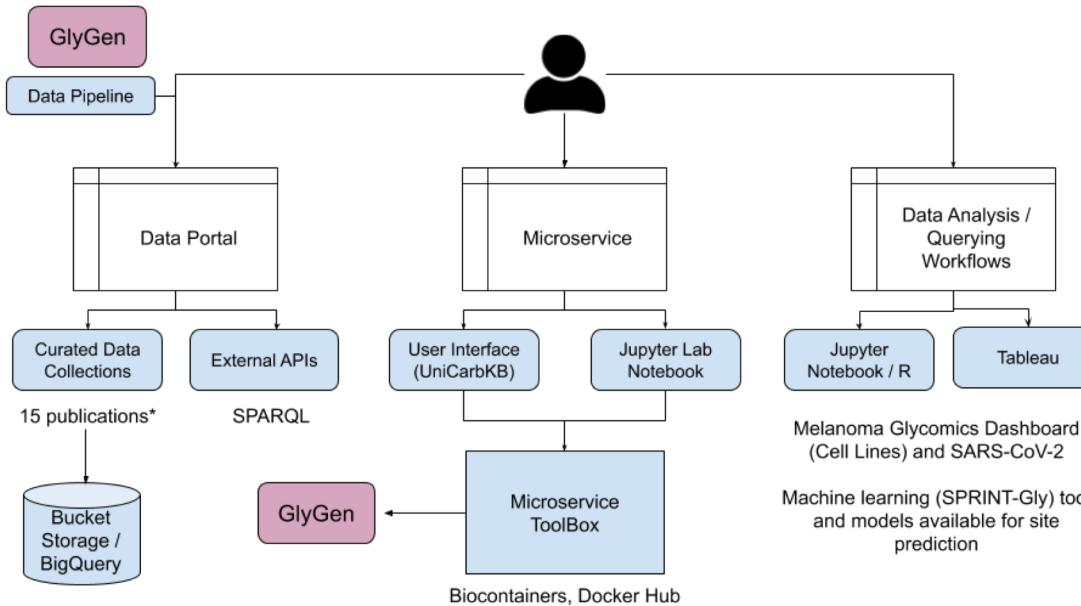
Created a draft ontological representation of ~8200 UniCarbKB glycosylated proteoforms (under review by GlyGen team)

FUTURE: Disseminate relevant PRO terms and annotations via PRO and GlyGen web portals



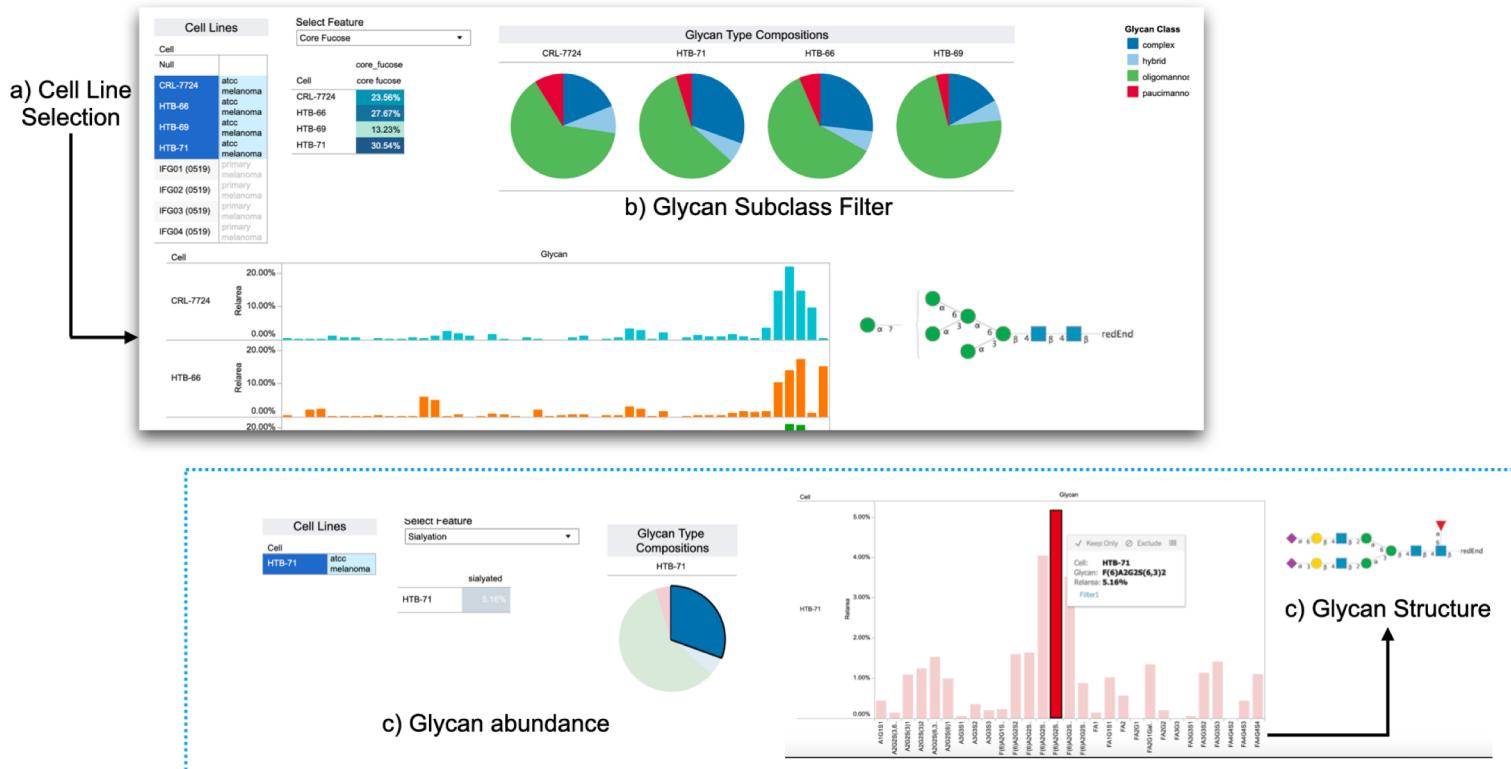
Darren Natale
Karen Ross

UniCarbKB (Griffith University)

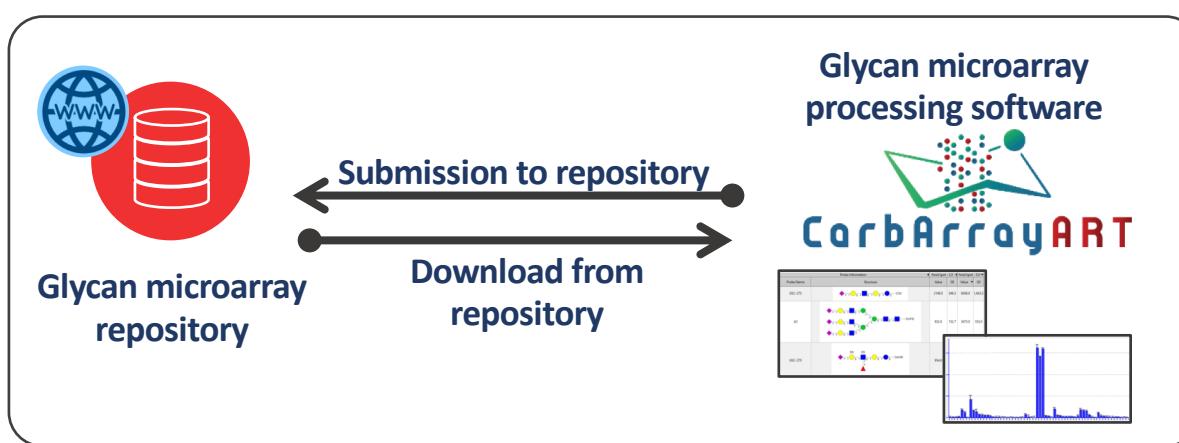


Matthew Campbell
Jodie Abrahams

- Continued biocuration (Human, Mouse and Rat) high-throughput and purified glycoprotein (site and global analysis)
- Addition and expansion of HCV and SARS-CoV
- Continued support for international data standardisation activities via SPARQL
- Curation of data that adheres to MIRAGE guidelines and adoption of ontologies.



Data visualization tools for users to query and compare curated GlyGen data collections. Current releases designed for cancer cell lines and SARS-CoV-2



Ten Feizi
Yukie Akune

CarbArrayART (**C**arbohydrate **m**icro**A**rray **A**nalysis and **R**eporting **T**ool)

Design of new systems for semi-automatic data transfer package to include

- Array geometry (locations of printed probes)
- Raw data (scan data and images)
- Processed data including average binding intensities after filtering and sorting of glycan probes (for tabulations, bar charts and heatmap presentations)
- Metadata (MIRAGE compliant)

GLycomics Array Dashboard (GLAD)

- A glycan microarray data visualization and analysis tool.
- Enables simultaneous visualization of glycan structure alongside microarray data. <https://glycotoolkit.com/GLAD/>



Rick Cummings
Akul Mehta

GlyMDB: Glycan Microarray Database

- Web-based tool which correlates glycan microarray data and shows corresponding PDB structures for the glycan binding proteins tested on the microarray.
- Protein sequence-based search using BLAST matches microarray data to PDB structures containing glycans. <http://www.glycanstructure.org/glymdb>

GlycoGlyph

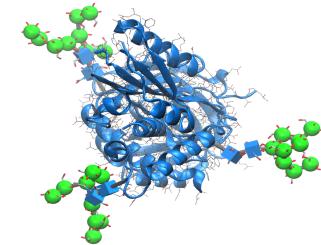
- An open source web-based glycan drawing and naming tool.
- Helps draw glycan structures in SNFG format using a graphical user interface or CFG linear nomenclature. <https://glycotoolkit.com/glycoglyph>

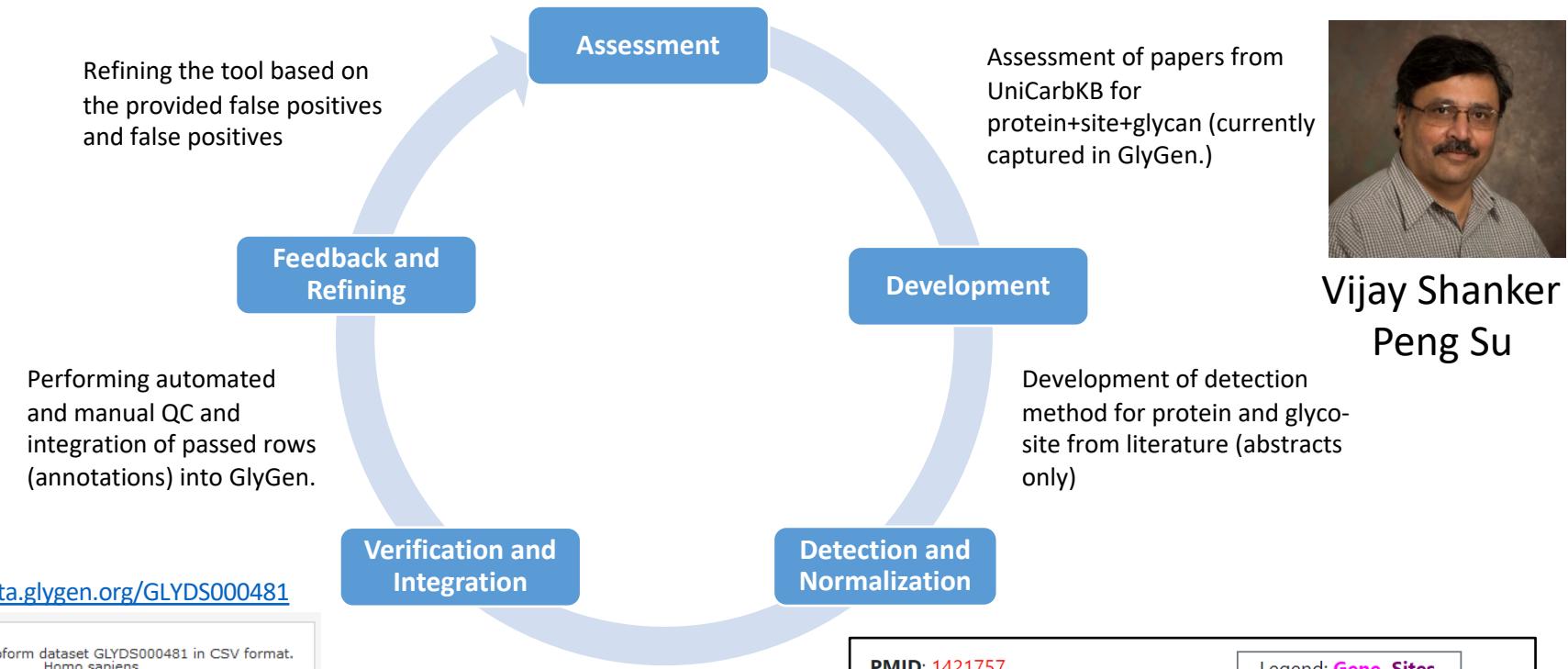
- ▶ Improved performance of glycoprotein builder
 - ▶ Currently 0.5 secs (simple structures) to 1-2 hrs (HIV env)
 - ▶ Code optimization and parallelization (CPU or GPU)
- ▶ Fully Implemented JSON API
 - ▶ Enabled access to carbohydrate builder and glycoprotein builder
 - ▶ Input and Output in JSON – easy format
 - ▶ Usage Examples:<https://github.com/GLYCAMS-Web/website/tree/master/Examples/JsonApi>
- ▶ Implement a robust PDB pre-processor
- ▶ JSON API would enable GlyGen to download 3D co-ordinates to be able to show it on the front end



Rob Woods

Glycoprotein with Man9 glycan





<https://data.glygen.org/GLYDS000481>

Reviewed proteoform dataset GLYDS000481 in CSV format.
Homo sapiens

Human Glycosylation Sites [Automatic Literature Mining]

uniprotkb_ca...	glycosylatio...
Q6PIU2-1	270...
Q6PIU2-1	270...

List of human [taxid:9606] proteins with information on glycosylation sites detected from PubMed dat ...
[view details](#)

Additional output: Glycan list/dictionary

Integration of the detected and verified annotations in GlyGen.

Assessment of papers from UniCarbKB for protein+site+glycan (currently captured in GlyGen.)



Vijay Shanker
Peng Su

PMID: 1421757

Legend: Gene, Sites

Title : Identification of the O-linked glycosylation **site** of the human **transferrin receptor**

Abstract :

1. The human **transferrin receptor** is a **glycoprotein** containing three N-linked and one O-linked glycosylation **sites**
2. Tryptic digestion of the **receptor**, followed by chromatography on BioGel P-2 and reverse-phase HPLC, yields a **glycopeptide** (amino acids 101-120) containing the O-linked **site**
3. Amino acid sequence analysis reveals that the **site** of O-glycosylation is **Thr-104**

Detection of protein and site.

GlyGen

Progress in Portal Enhancement

The Home Page

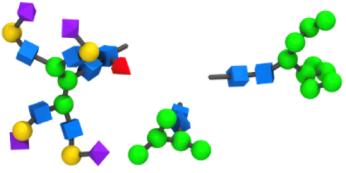
Demo Session B: “GlyGen Features Demonstration” for introduction to functionalities

GlyGen  [HOME](#) [EXPLORE](#) [QUICK SEARCH](#) [TRY ME](#) [DATA](#) [HELP](#) [MORE](#) [Beta Testing](#) [MY GLYGEN](#)



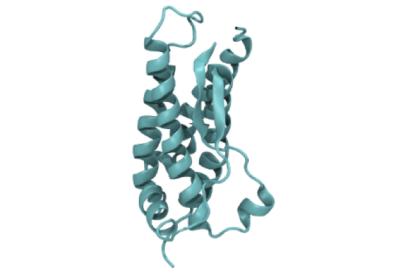
GlyGen
Computational and Informatics Resources for Glycoscience

GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.



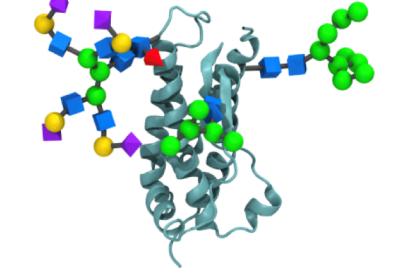
Glycan
Search for glycan structures based on their chemical and structural properties.

[EXPLORE](#)



Protein
Search for proteins based on their sequences, accessions, and annotations.

[EXPLORE](#)



Glycoprotein
Search for glycoproteins based on the protein or glycan structures involved.

[EXPLORE](#)

Version

Portal: 1.5 (24/Apr/2020)
Webservice: 1.5.33 (04/May/2020)
Data: 1.5.23 (24/Apr/2020)

Your Opinion Matters



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

[LEAVE FEEDBACK](#)

Database Statistics

Homo sapiens	
Glycans	4080
Proteins	20656
Glycoproteins	5241
Mus musculus	
Glycans	597
Proteins	21959
Glycoproteins	3884
Rattus norvegicus	
Glycans	290
Proteins	21647
Glycoproteins	2167

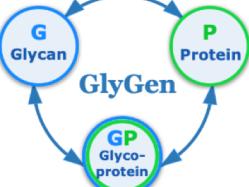
Highlights of Portal Improvements in Year 3

Rapid versioning cycle; every 3 months during year 3, v1.3-v1.6

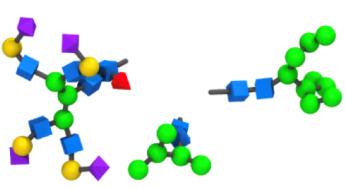
GlyGen  [HOME](#) [EXPLORE](#) [QUICK SEARCH](#) [TRY ME](#) [DATA](#) [HELP](#) [MORE](#) [Beta Testing](#) [MY GLYGEN](#)

GlyGen
Computational and Informatics Resources for Glycoscience

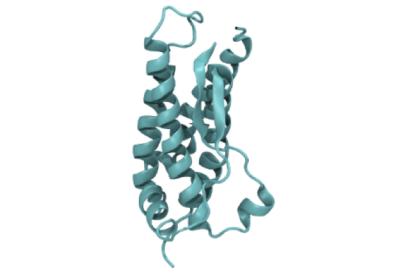
GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.



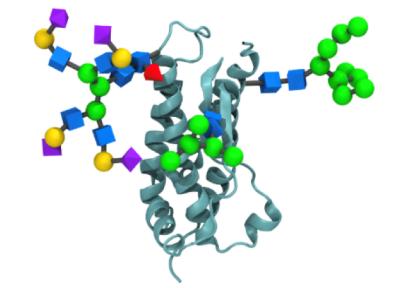
Glycan
Search for glycan structures based on their chemical and structural properties.
[EXPLORE](#)



Protein
Search for proteins based on their sequences, accessions, and annotations.
[EXPLORE](#)



Glycoprotein
Search for glycoproteins based on the protein or glycan structures involved.
[EXPLORE](#)



Version

Portal: 1.5 (24/Apr/2020)
Webservice: 1.5.33 (04/May/2020)
Data: 1.5.23 (24/Apr/2020)

Your Opinion Matters



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

[LEAVE FEEDBACK](#)

Database Statistics

Homo sapiens	
Glycans	4080
Proteins	20656
Glycoproteins	5241
Mus musculus	
Glycans	597
Proteins	21959
Glycoproteins	3884
Rattus norvegicus	
Glycans	290
Proteins	21647
Glycoproteins	2167

Highlights of Portal Improvements in Year 3

Enhanced Global Search



GlyGen

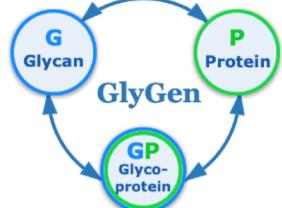
HOME EXPLORE QUICK SEARCH TRY ME DATA HELP MORE HGF



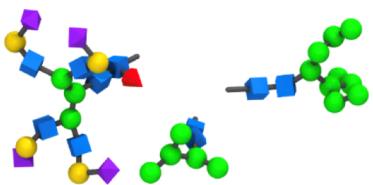
Beta Testing MY GLYGEN

GlyGen

Computational and Informatics Resources for Glycoscience



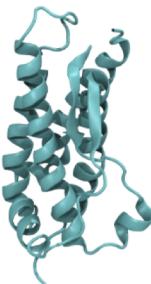
GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.



Glycan

Search for glycan structures based on their chemical and structural properties.

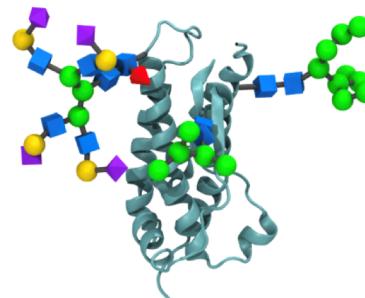
[EXPLORE](#)



Protein

Search for proteins based on their sequences, accessions, and annotations.

[EXPLORE](#)



Glycoprotein

Search for glycoproteins based on the protein or glycan structures involved.

[EXPLORE](#)

Version

Portal: 1.5 (08/Jun/2020)
Webservice: 1.5.36 (12/May/2020)
Data: 1.5.28 (08/Jun/2020)

Your Opinion Matters



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

[LEAVE FEEDBACK](#)

Database Statistics

Homo sapiens

Glycans	915
Proteins	20656
Glycoproteins	5245

Mus musculus

Glycans	92
Proteins	21959
Glycoproteins	3883

Rattus norvegicus

Glycans	167
Proteins	21647
Glycoproteins	2106

More Intuitive Search Results for Global Search

Search result for HGF

11 Glycans
match "HGF"

in Identifiers	0
in Subtype	0
in Enzyme	0
in Motifs	0
in Sequences	0
in Glycoproteins	11
in Type	0
in Species	0

117 Proteins
match "HGF"

in Gene	20
in Function	88
in Disease	0
in Identifiers	9
in Names	0
in Species	0

61 Glycoproteins
match "HGF"

in Gene	14
in Function	40
in Disease	0
in Identifiers	7
in Names	0
in Species	0

Feedback

Explore this

Feedback Widget Allows Users to Send Feedback from Any Page

Search result for HGF

11 Glycans	
match "HGF"	
in Identifiers	0
in Subtype	0
in Enzyme	0
in Motifs	0
in Sequences	0
in Glycoproteins	11
in Type	0
in Species	0

117 Proteins	
match "HGF"	
in Gene	20
in Function	88
in Disease	0
in Identifiers	9
in Names	0
in Species	0

61 Glycoproteins	
match "HGF"	
in Gene	14
in Function	0
in Disease	0
in Identifiers	7
in Names	0
in Species	0

Feedback

What would you like to share with us?

Problem Question Suggestion

Your feedback

Your Name (will not be published)

Your Email (will not be published)

More Intuitive Search Results for Global Search

Search result for HGF

11 Glycans
match "HGF"

in Identifiers	0
in Subtype	0
in Enzyme	0
in Motifs	0
in Sequences	0
in Glycoproteins	11
in Type	0
in Species	0

117 Proteins
match "HGF"

in Gene	20
in Function	88
in Disease	0
in Identifiers	9
in Names	0
in Species	0

61 Glycoproteins
match "HGF"

in Gene	14
in Function	40
in Disease	0
in Identifiers	7
in Names	0
in Species	0

Feedback

Explore this

Glycoprotein Search Result for HGF

General Search / Glycoprotein List

Summary of your Glycoprotein Search

Performed on: June 12th 2020, 9:19:47 pm (EST)

Search Term: HGF

[Update Results](#)[Modify Search](#)** To perform the same search again using the current version of the database, click "[Update Results](#)".

Page

Prev

1

Next

Records per page

20

"14 Proteins were found"

 [DOWNLOAD](#) ▾

Feedback

UniProtKB Accession	Gene Name	UniProtKB Name	Chemical Mass (Da)	Organism	RefSeq Name	RefSeq Accession
P17945-1	Hgf; HPTA	Hepatocyte growth factor; Scatter factor; SF; Hepatopoietin-A	82906	Rattus norvegicus	hepatocyte growth factor preproprotein	NP_058713.1
Q9R098-1	HGFA; Hgfac	Hepatocyte growth factor activator; HGF activator; 3.4.21.-; HGFA	70568	Mus musculus	hepatocyte growth factor activator preproprotein	NP_062320.2
P08581-1	MET; AUTS9; c-Met; DFNB97; HGFR; RCCP2	Hepatocyte growth factor receptor; HGF receptor; 2.7.10.1; Scatter factor receptor; SF receptor; Proto-oncogene c-Met; Tyrosine-protein kinase Met	155541	Homo sapiens	hepatocyte growth factor receptor isoform b preproprotein	NP_000236.2
Q96FE7-1	HGFL; HGFL; hHGFL(S); TrIP; PIK3IP1	Phosphoinositide-3-kinase-interacting protein 1; Kringle domain-containing protein HGFL	28248	Homo sapiens	phosphoinositide-3-kinase-interacting protein 1 isoform 2 precursor	NP_001129383.1
P14210-1	HGF; HPTA; DFNB39; F-TCF; HGFB; HPTA; SF	Hepatocyte growth factor; Scatter factor; SF; Hepatopoietin-A	83134	Homo sapiens	hepatocyte growth factor isoform 1 preproprotein	NP_000592.3
Q08048-1	Hgf; C230052L06Rik; HGF/SF; NK1; NK2; SF; SF/HGF	Hepatocyte growth factor; Scatter factor; SF; Hepatopoietin-A	82945	Mus musculus	hepatocyte growth factor isoform 1	NP_001276387.1

[Explore this](#)

Glycoprotein Detail Page for human HGF

General

Species

Function

GO Annotation

Glycosylation

Sequence

Pathway

Isoforms

Homologs

Disease

Mutation

Expression Tissue

Expression Disease

Cross References

Publications

Explore this

General Search / Glycoprotein List / Glycoprotein Detail

Details for Glycoprotein P14210-1

DOWNLOAD ▾

General

- Gene Name: HGF; HPTA; DFNB39; F-TCF; HGFB; HPTA; SF
- Gene Location: Chromosome: 7 (81,770,438 - 81,699,008)

Ensembl Gene 1

- UniProtKB ID: HGF_HUMAN
- UniProtKB Accession: P14210-1
- UniProtKB Accession Length: 728
- UniProtKB Entry Name: Hepatocyte growth factor; Scatter factor; SF; Hepatopoietin-A
- Chemical Mass: 83,134 Da
- RefSeq Accession: NP_000592.3
- RefSeq Name: hepatocyte growth factor isoform 1 preproprotein
- RefSeq Summary: This gene encodes a protein that binds to the hepatocyte growth factor receptor to regulate cell growth, cell motility and morphogenesis in numerous cell and tissue types. Alternative splicing results in multiple transcript variants, at least one of which encodes a preproprotein that is proteolytically processed to generate alpha and beta chains, which form the mature heterodimer. This protein is secreted by mesenchymal cells and acts as a multi-functional cytokine on cells of mainly epithelial origin. This protein also plays a role in angiogenesis, tumorogenesis, and tissue regeneration. Although the encoded protein is a member of the peptidase S1 family of serine proteases, it lacks peptidase activity. Mutations in this gene are associated with nonsyndromic hearing loss. [provided by RefSeq, Nov 2015].

Incorporation of RefSeq details

Species

- Homo sapiens UniProtKB 1
- Taxonomy ID: 9606

Site-Specific Glycosylation Information

Explore this

Glycosylation

With Reported Glycans Without Reported Glycans

Sources	GlyTouCan Accession	Type	Residue	Image of Glycan Structure
PubMed (2) UniCarbKB (1)	G60542VK	N-linked	Asn294	
PubMed (1) GlyConnect (1)	G60542VK	N-linked	Asn294	
PubMed (1) GlyConnect (1)	G17689DH	N-linked	Asn294	
PubMed (1) GlyConnect (1)	G74239ZQ	N-linked	Asn294	

Feedback

Sequence Context for Glycosylation Site



General

Sequence view

Site Annotation

Protein Search / Protein List / Protein Detail / ProtVista View / Site View

Site view P14210-1 294

 [DOWNLOAD ▾](#)

General

- **Gene Name:** HGF; HPTA; DFNB39; F-TCF; HGFB; HPTA; SF
- **UniProtKB ID:** HGF_HUMAN
- **UniProtKB Accession:** P14210-1
- **UniProtKB Entry Name:** Hepatocyte growth factor; Scatter factor; SF; Hepatopoietin-A
- **RefSeq Accession:** NP_000592.3
- **RefSeq Name:** hepatocyte growth factor isoform 1 preproprotein
- **Length:** 728 aa
- **Monoisotopic Mass:** 83134 Da

Sequence View

N-294

294-Asn

<< E Y C A I K T C A D N T M N D T D V P L E T >>

Site Annotation

Annotation	Position	Source	Additional Information
N-Glycosylation	294	UniProtKB 1	NA
N-Glycosylation	294	PubMed 2	/

Scroll Down



Glycan Heterogeneity at Individual Sites

General

Sequence view

Site Annotation

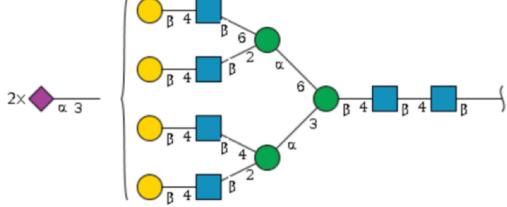
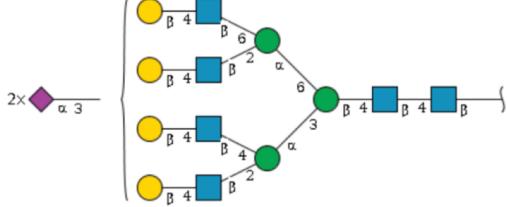
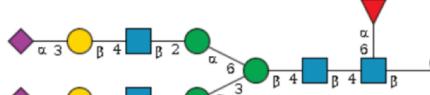
Sequence View

N-294

294-Asn

<< E Y C A I K T C A D N T M N D T D V P L E T >>

Site Annotation

Annotation	Position	Source	Additional Information
N-Glycosylation	294	UniProtKB 1	NA
N-Glycosylation	294	PubMed 2 UniCarbKB 1	
N-Glycosylation	294	PubMed 1 GlyConnect 1	
N-Glycosylation	294	PubMed 1 GlyConnect 1	

Feedback

Back to Glycoprotein Detail Page. Sequence Views

Explore this

- General
- Species
- Function
- GO Annotation
- Glycosylation
- Sequence**
- Pathway
- Isoforms
- Homologs
- Disease
- Mutation
- Expression Tissue
- Expression Disease
- Cross References
- Publications

Sequence

ProtVista 

+10 +20 +30 +40 +50

1 MWTKLLPAL LLQHVLHLI LPIIAIPYAE GQRKRRNTIH EFKKSAKTTL IKIDPALKIK
61 TKKVNTADQC ANRCTRNKGL PFTCKAFVFD KARKOCLWFP FNSMSSGVKK EFGHEFDLYE
121 NKDYIRNCII GKGRSYKGTV SITKSGIKCQ PWSSMIPHEH SFLPSSYRGK DLQENYCRNP
181 RGEEGGPWCFC TSNPEVRYEV CDIPQCSEVE CMTONGESYR GLMDHTESGK ICQRWDHQTP
241 HRHKFLPERY PDKGFDNNYC RNPDGQPRPW CYTLDPHTRW EYCAIKTCAD NTMNDTDVPL
301 ETTECIQGQG EGYRGTVTNTI WNGIPCQRWD SQYPHEHDMT PENFKKDRL ENYCRNPDGS
361 ESPWCFTTDP NIRVGYCSQI PNCDMSHQQD CYRGNKNYM GNLSQTRSGL TCSMWDKNME
421 DLHRHIFWEP DASKLNENYC RNPPDDDAHGP WCYTGPNLIP WDYCPISRCE GDTTPTIVNL
481 DHPVISCAKT KQLRVVNGIP TRTNIGWMVS LRYRNKHICG GSLIKESWVL TARQCFPSRD
541 LKDYEAWLGI HDVHGRGDEK CKQVLNVSQL VYGPEGSDLV LMKLARPABL DDFVSTIDL
601 NYGCTIPEKT SCSVYGWGYT GLINYDGLLR VAHLYIMGNE KCSQHHRGKV TLNESEICAG
661 AEKIGSGPCE GDYGGPLVCE QHKMRMVLGV IVPGRGCAIP NRPGIFVRVA YYAKWIHKII
721 LTYKVPQS

N-linked Sites 4

O-linked Sites 1

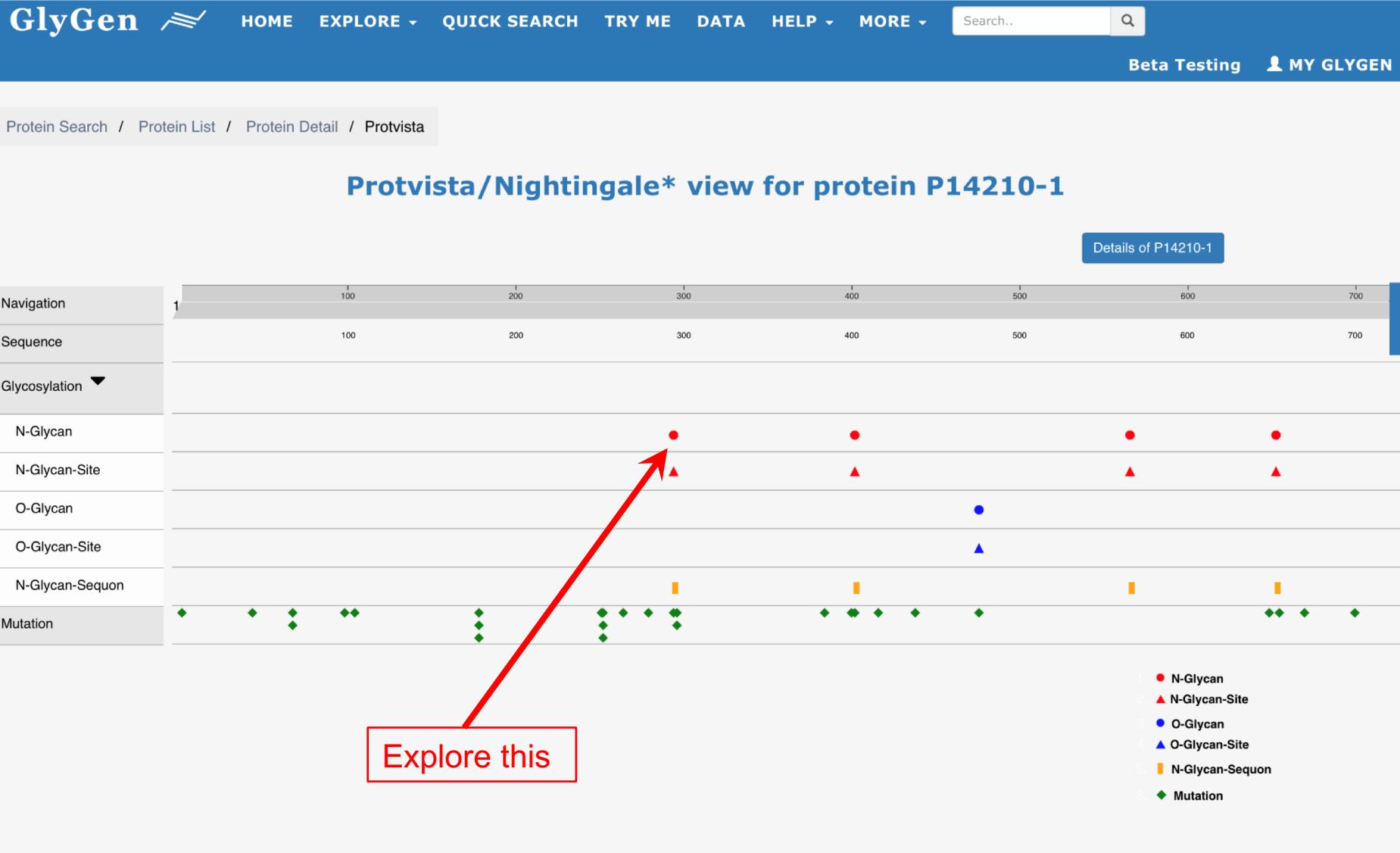
Mutation 22

Sequon 4

Pathway

- **Reactome:**
 - Interleukin-7 signaling [R-HSA-1266695](#)
 - Negative regulation of MET activity [R-HSA-6807004](#)
 - MET activates STAT3 [R-HSA-8875791](#)
 - Interleukin-4 and Interleukin-13 signaling [R-HSA-6785807](#)
 - PIP3 activates AKT signaling [R-HSA-1257604](#)
 - MET interacts with TNS proteins [R-HSA-8875513](#)
 - MET activates PTK2 signaling [R-HSA-8874081](#)
 - MET activates RAS signaling [R-HSA-8851805](#)
 - MET activates RAP1 and RAC1 [R-HSA-8875555](#)
 - MET activates PTPN11 [R-HSA-8865999](#)
 - MET activates PI3K/AKT signaling [R-HSA-8851907](#)
 - RAF/MAP kinase cascade [R-HSA-5673001](#)
 - PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling [R-HSA-6811558](#)
 - MET Receptor Activation [R-HSA-6806942](#)
 - Constitutive Signaling by Aberrant PI3K in Cancer [R-HSA-2219530](#)
 - MET receptor recycling [R-HSA-8875656](#)
 - Platelet degranulation [R-HSA-114608](#)
- **KEGG Pathway:**
 - [hsa:3082](#)

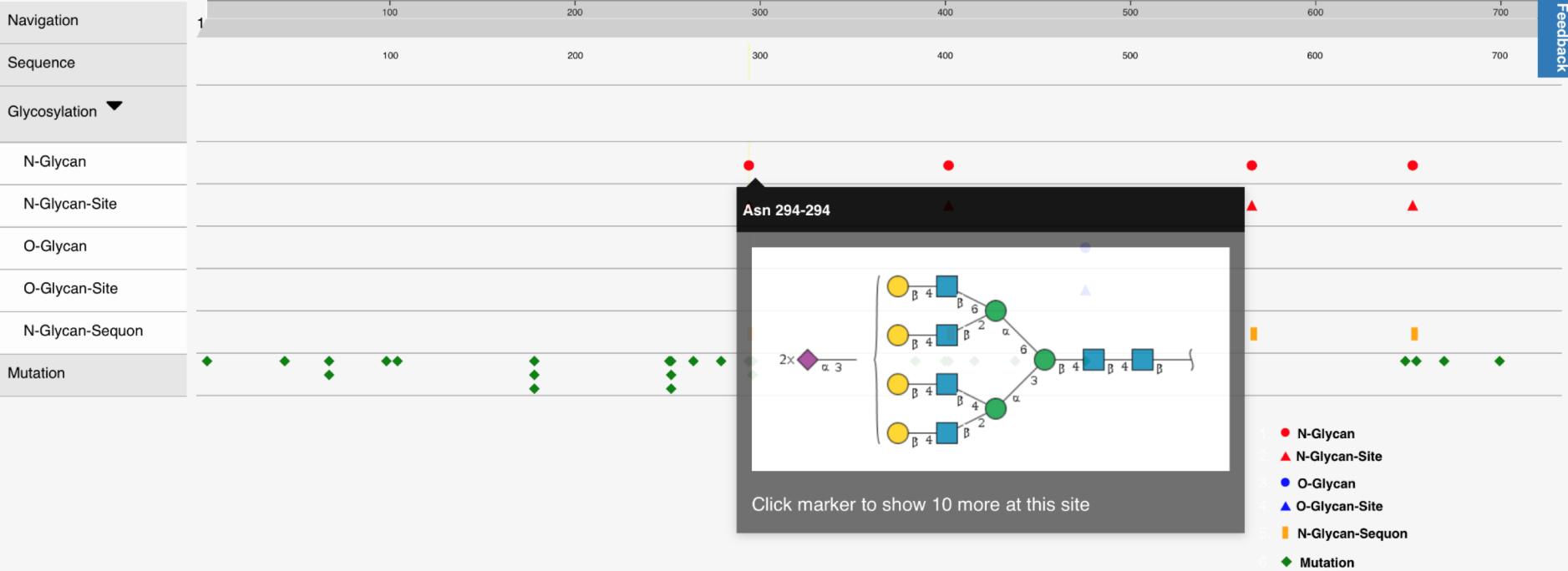
ProtVista Provides Summary of Sequence Features



Hover Over Site Opens Site-Specific Data



Protvista/Nightingale* view for protein P14210-1

[Details of P14210-1](#)[Feedback](#)

Visualization of Glycosylation Sites, Consensus, and Mutations on Canonical Sequence

- General
- Species
- Function
- GO Annotation
- Glycosylation
- Sequence**
- Pathway
- Isoforms
- Homologs**
- Disease
- Mutation
- Expression Tissue
- Expression Disease
- Cross References
- Publications

Explore this

Sequence

+10 +20 +30 +40 +50

1 MWVTKLLPAL LLQHVLHLI LLPIAIPYAE GQRKRRNTIH EFK**K**SAKTTL IKIDPALKIK
61 TKKVNT**A**DQC ANRCTRNLGK PFTCAFVFD KARKQCL**W**FP FNMS**S**SGVKK EFGHEFDLYE
121 NKDYIRNCII GKGRSYKGTV SITKSGIKCQ PWSSMIPHEH SFLPSSYRGK DLQENYC**R**NP
181 RGEEGGPWCF TSMPEVRYEV CDIPQCSEVE CMTONGESYR GLMDHTESGK ICQRWDHQTP
241 HRHKFLPERY **P**DKGFDNYC RNPN**G**QPRPW CYTLDPT**R**W EYCAIKTCAD NTM**N**DTDVPL
301 ETTECIQGQG EGYRGTVNTI WNGIIPCQRWD SQYPHEHDMDT PENFKCKDLR ENYCRNPDGS
361 ESPWCFTTDP NIRVGYSQI PNCPMSHGQD CYRGNKG**N**Y **G**NLSQTRSGL TCSMW**K**NME
421 DLHRHIFWEP DASKLN**E**NYC RNPPDDAHGP WCYTGNPLIP WDYCPISRCE GDTPPTIVNL
481 DHPVISCAKT KQLRVVNGIP TRTNIGWMVS LRYRNKHICG GSLIKESWVL TARQCFPSRD
541 LKDYEAWLGI HDVHGRGDEK CKQVLN**V**NSQL VYGPEGSDLV LMKLARPALV DDFVSTIDLP
601 NYGCTIPEKT SCSYVGWGYT GLINYDGLLR VAHLYIMGNE KCSQHHRG**K**V TL**N**ESIECAG
661 AEKIGSGP**C**E GDYGGPLVCE QHKMRMVLGV IVPGRGCA**P** NRPGIFVRVA YYAKWIHKII
721 LTYKVPQS

ProtVista

N-linked Sites (4)

O-linked Sites (1)

Mutation (22)

Sequon (4)

Toggle visualization of sites, consensus, and identified mutations

Pathway

- **Reactome:**
 - Interleukin-7 signaling [R-HSA-1266695](#)
 - Negative regulation of MET activity [R-HSA-6807004](#)
 - MET activates STAT3 [R-HSA-8875791](#)
 - Interleukin-4 and Interleukin-13 signaling [R-HSA-6785807](#)
 - PIP3 activates AKT signaling [R-HSA-1257604](#)
 - MET interacts with TNS proteins [R-HSA-8875513](#)
 - MET activates PTK2 signaling [R-HSA-8874081](#)
 - MET activates RAS signaling [R-HSA-8851805](#)
 - MET activates RAP1 and RAC1 [R-HSA-8875555](#)
 - MET activates PTPN11 [R-HSA-8865999](#)
 - MET activates PI3K/AKT signaling [R-HSA-8851907](#)
 - RAF/MAP kinase cascade [R-HSA-5673001](#)
 - PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling [R-HSA-6811558](#)
 - MET Receptor Activation [R-HSA-6806942](#)
 - Constitutive Signaling by Aberrant PI3K in Cancer [R-HSA-2219530](#)
 - MET receptor recycling [R-HSA-8875656](#)
 - Platelet degranulation [R-HSA-114608](#)
- **KEGG Pathway:**
 - [hsa:3082](#)

Explore Glycoprotein Isoforms

General
Species
Function
GO Annotation
Glycosylation
Sequence
Pathway
Isoforms
Homologs
Disease
Mutation
Expression Tissue
Expression Disease
Cross References
Publications

Isoforms

Alignment

Show Sequences

Explore this

- **UniProtKB Isoform Accession:** [C9WSJ3-1](#)
Chromosome: 7 (81,769,971 - 81,755,994)
UniProtKB Accession Length: 184

Ensembl Transcript 1

Ensembl Peptide 1

- **UniProtKB Isoform Accession:** [C9JDP4-1](#)
Chromosome: 7 (81,769,971 - 81,757,248)
UniProtKB Accession Length: 141

Ensembl Transcript 1

Ensembl Peptide 1

- **UniProtKB Isoform Accession:** [AOA2R8YEI1-1](#)
Chromosome: 7 (81,769,971 - 81,752,115)
UniProtKB Accession Length: 205

Ensembl Transcript 1

Ensembl Peptide 1

- **UniProtKB Isoform Accession:** [C9WSJ4-1](#)
Chromosome: 7 (81,769,971 - 81,757,188)
UniProtKB Accession Length: 161

Ensembl Transcript 1

Ensembl Peptide 1

- **UniProtKB Isoform Accession:** [P14210-1](#)
Chromosome: 7 (81,769,971 - 81,702,584)
UniProtKB Accession Length: 728

Ensembl Transcript 1

Ensembl Peptide 1

Feedback

Align Glycoprotein Isoforms (can also be done for homologs)



Protein Search / Protein List / Protein Detail / Isoform
Alignment

Isoform Alignment for P14210-1

Alignment

9606 Homo sapiens P14210-3	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens P14210-2	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens P14210-1	HGF_HUMAN	Hepatocyte growth factor	MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens P14210-6	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens P14210-5	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens P14210-4	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens C9WSJ4-1	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens C9WSJ3-1	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens A0A2R8YEI1-1	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens C9JS80-1	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
9606 Homo sapiens C9JDP4-1	HGF_HUMAN		MWVTKLLPALLLQHVLLHLLLLPIAIPYAEQQRKRRTIHEFKSAKTTLIKIDPALKIK
<hr/>			
9606 Homo sapiens P14210-3	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens P14210-2	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens P14210-1	HGF_HUMAN	Hepatocyte growth factor	TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens P14210-6	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens P14210-5	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens P14210-4	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens C9WSJ4-1	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens C9WSJ3-1	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens A0A2R8YEI1-1	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens C9JS80-1	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
9606 Homo sapiens C9JDP4-1	HGF_HUMAN		TKKVNNTADQCANRCTRNKGLPFTCKAFVFDKARKQCLWFPFNMSGVKKEFGHEFDLYE
<hr/>			
9606 Homo sapiens P14210-3	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEH-----SYRGKDLQENYCRNP
9606 Homo sapiens P14210-2	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEHSFLPSSYRGKDLQENYCRNP
9606 Homo sapiens P14210-1	HGF_HUMAN	Hepatocyte growth factor	NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEHSFLPSSYRGKDLQENYCRNP
9606 Homo sapiens P14210-6	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEHSFLPSSYRGKDLQENYCRNP
9606 Homo sapiens P14210-5	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEH-----SYRGKDLQENYCRNP
9606 Homo sapiens P14210-4	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEHSFLPSSYRGKDLQENYCRNP
9606 Homo sapiens C9WSJ4-1	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEH-----
9606 Homo sapiens C9WSJ3-1	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEH-----
9606 Homo sapiens A0A2R8YEI1-1	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEH-----
9606 Homo sapiens C9JS80-1	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEH-----
9606 Homo sapiens C9JDP4-1	HGF_HUMAN		NKDYIRNCIIGKGRSYKGTVSITKSGIKCQPWSSMIPHEH-----

Back to Homepage, Enhanced Search Features

Choose Search Type

GlyGen

HOME

EXPLORE ▾

QUICK SEARCH

TRY ME

DATA

HELP ▾

MORE ▾

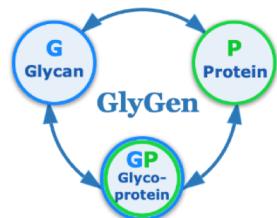
Search..



Beta Testing

MY GLYGEN

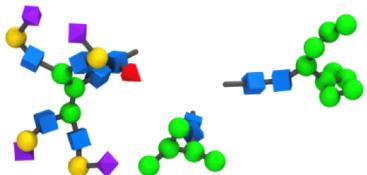
Glycan Search
Protein Search
Glycoprotein Search



GlyGen

Computational and Informatics Resources for Glycoscience

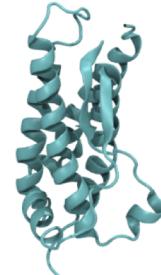
GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.



Glycan

Search for glycan structures based on their chemical and structural properties.

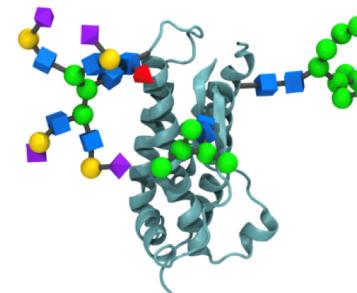
EXPLORE



Protein

Search for proteins based on their sequences, accessions, and annotations.

EXPLORE



Glycoprotein

Search for glycoproteins based on the protein or glycan structures involved.

EXPLORE



Please provide feedback and suggestions to help us improve the GlyGen portal and make it more useful for the community.

LEAVE FEEDBACK

Version

Portal: 1.5 (08/Jun/2020)
Webservice: 1.5.36 (12/May/2020)
Data: 1.5.28 (08/Jun/2020)

Your Opinion Matters

Database Statistics

Homo sapiens

Glycans	915
Proteins	20656
Glycoproteins	5245

Mus musculus

Glycans	92
Proteins	21959
Glycoproteins	3883

Rattus norvegicus

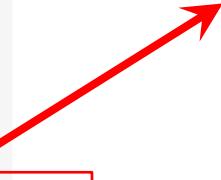
Glycans	167
Proteins	21647

Improved Glycan Search Features

GlyGen  HOME EXPLORE  QUICK SEARCH TRY ME DATA HELP  MORE  Search..  Beta Testing  MY GLYGEN

Glycan Search

Simple Search **Advanced Search** Composition Search Tutorial

Explore this 

Any category 

Enter the search term

Examples: G17689DH, P14210, B4GALT1, Homo sapiens

Search

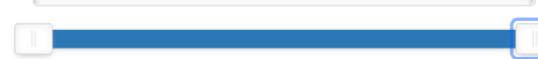
* **Any category** allows you to search an entire GlyGen database, including the context match. *

Advanced Search Features

Glycan Search

[Simple Search](#)[Advanced Search](#)[Composition Search](#)[Tutorial](#)[Clear Fields](#)[Search Glycan](#) **ⓘ Glycan ID:** Enter single or multiple comma-separated GlyTouCan Accession(s) or Cross Reference(s) ID.GlyTouCan Accession Example: [G17689DH](#)Explore [GlyTouCan Accession](#)Cross References ID Example: [G10716](#) **ⓘ Monoisotopic Mass:**

Native



Min 149 Max 6,752

 ⓘ Number of Sugars:

Min -1 Max 38

 ⓘ Organism: Click to select multiple Organisms

OR

 **ⓘ Glycan Type:** Select Glycan Type **ⓘ Glycosylated Protein:** Enter the UniProtKB Accession of your proteinExample: [P14210](#)Explore [UniProtKB Accession](#) **ⓘ Glycan Motif:** Enter the name of a Glycan Motif contained in your glycanExample: [N-Glycan complex](#)Explore [Glycan Motif](#)**Select Native or PerMe Mass**

Additional Search Parameters in Glycan, Protein, and Glycoprotein Searches



Glycan Search

[Simple Search](#)[Advanced Search](#)[Composition Search](#)[Tutorial](#)[Clear Fields](#)[Search Glycan](#)**Glycan ID:** Enter single or multiple comma-separated GlyTouCan Accession(s) or Cross Reference(s) ID.

GlyTouCan Accession Example: [G17689DH](#) Explore GlyTouCan Accession
Cross References ID Example: [G10716](#)

Monoisotopic Mass:

Permethylated
✓ Native

Min 1,637 Max 5,129

Number of Sugars:

Min -1 Max 38

Organism: Click to select multiple Organisms

OR

Glycan Type: Select Glycan Type**Glycosylated Protein:** Enter the UniProtKB Accession of your protein
Example: [P14210](#) Explore UniProtKB Accession**Glycan Motif:** Enter the name of a Glycan Motif contained in your glycan
Example: [N-Glycan complex](#) Explore Glycan Motif

User Assists Added Across All Search Types

Glycan Search

Simple Search

Advanced Search

Composition Search

Tutorial

Clear Fields

Search Glycan

Tool tips for parameters in all types of searches appear by hovering over “?”

ⓘ Glycan ID:

ⓘ Monoisotopic Mass:

The monoisotopic mass is the sum of the masses of the atoms in a molecule.

Use the sliders to select a Monoisotopic Mass range for your protein(s)

Permethylated



Min 1,637

Max 5,129



Min -1

Max 38

ⓘ Organism:

Click to select multiple Organisms

OR

ⓘ Glycan Type:

Select Glycan Type

ⓘ Glycosylated Protein:

Enter the UniProtKB Accession of your protein

Example: P14210

Explore UniProtKB Accession

ⓘ Glycan Motif:

Enter the name of a Glycan Motif contained in your glycan

Example: N-Glycan complex

Explore Glycan Motif

Added Glycan Composition Search

Glycan Search

Simple Search

Advanced Search

Composition Search

Tutorial

Explore this

Feedback

Residue

Hexose:*Hex, e.g., Glc, Gal, Man***N-Acetyl-Hexosamine:***HexNAc, e.g., GlcNAc, GalNAc***Hexosamine:***HexN, e.g., GlcN***Deoxy-Hexose:***dHex, e.g., Fuc***N-Acetyl-Neuraminic acid:***Sialic acid as NeuAc***N-Glycolyl-Neuraminic acid:***Sialic acid as NeuGc***Uronic acid:***UA, e.g., GlcA, GalA***Phosphate:***PO₃***Sulfate:***SO₄***Other:****Contains**

Maybe

Min

0

Max

15

0

14

0

2

0

7

0

5

0

5

0

6

0

2

0

5

0

6

All No

All Yes

Clear Fields

Undo

Redo

Search Glycan

Enter Composition from Analytic Data, such as LC-MS/MS

GlyGen  HOME EXPLORE  QUICK SEARCH TRY ME DATA HELP  MORE  Search..  Beta Testing  MY GLYGEN

Glycan Search

Simple Search Advanced Search **Composition Search** Tutorial

Residue	Contains	Min	Max
Hexose: <i>Hex, e.g., Glc, Gal, Man</i>	Yes 	5 	5 
N-Acetyl-Hexosamine: <i>HexNAc, e.g., GlcNAc, GalNAc</i>	Yes 	4 	4 
Hexosamine: <i>HexN, e.g., GlcN</i>	No 	0 	0 
Deoxy-Hexose: <i>dHex, e.g., Fuc</i>	Yes 	1 	1 
N-Acetyl-Neuraminic acid: <i>Sialic acid as NeuAc</i>	Yes 	2 	2 
N-Glycolyl-Neuraminic acid: <i>Sialic acid as NeuGc</i>	No 	0 	0 
Uronic acid: <i>UA, e.g., GlcA, GalA</i>	No 	0 	0 
Phosphate: <i>PO₃</i>	No 	0 	0 
Sulfate: <i>SO₃</i>	No 	0 	0 
Other:	No 	0 	0 

[All No](#) [All Yes](#) [Clear Fields](#) [Undo](#) [Redo](#) **Search Glycan**

Feedback

Returns Structural Possibilities Matching Compositions

Glycan Search / Glycan List

Links take user to additional information about glycan, including what other proteins carry this structure

Summary of your Glycan Search

Performed on: June 13th 2020, 1:11:21 pm (EST)

Hex:	5 - 5
HexNAC:	4 - 4
dHex:	1 - 1
NeuAc:	2 - 2

[Update Results](#)[Modify Search](#)

** To perform the same search again using the current version of the database, click "Update Results".

Page [Prev](#) [1](#) [Next](#) Records per page [20](#) "142 glycans were found" [DOWNLOAD ▾](#)

Feedback

GlyToCan Accession	Image of Glycan Structure	Monoisotopic Mass (Da)	Monoisotopic Mass-pMe (Da)	No of Sugars	On How Many Glycoproteins	How Many Enzymes
G88374WZ		2368.84	2943.48	12	80	0
G08146BT		2368.84	2943.48	12	16	0
G17689DH		2368.84	2943.48	12	11	29

Returns Structural Possibilities Matching Compositions

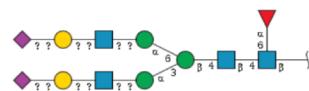
[General](#)[Species](#)[Motif](#)[Found Glycoproteins](#)[Biosynthetic Enzymes](#)[Digital Sequence](#)[Cross References](#)[Publications](#)

Glycan Search / Glycan List / Glycan Detail

Details for glycan G08146BT

[DOWNLOAD](#)

General



- GlyToCan Accession: G08146BT
- Monoisotopic Mass: 2,368.84 Da (Permethylated: 2943.48 Da)
- Composition: Hex₅ HexNAc₄ dHex₁ NeuAc₂
- Glycan Type/Subtype: N-glycan

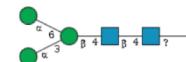
[Related glycans...](#)

From Glycan Detail page, can investigate related glycan structures

Species

Homo sapiens: [GlyConnect](#) (1) [UniCarbKB](#) (4)

Motif

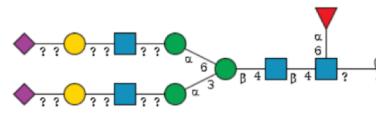


G00026MO-N-Glycan core basic

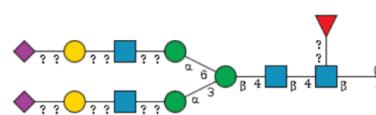
Found Glycoproteins

Mapping Informatic Relationships Between Glycans

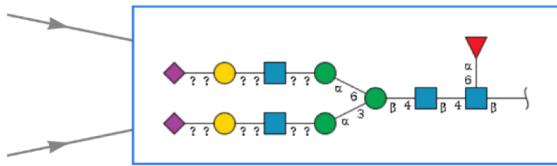
Demo Session C: “Glycan Ontology (GNOME) Substructure Searches and Motifs” for discussion of subsumption and new ways to represent the relationships between glycans



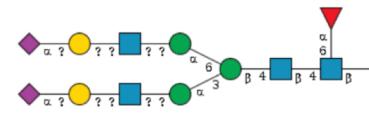
G00318AV



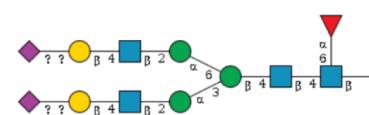
G72956NR



G08146BT



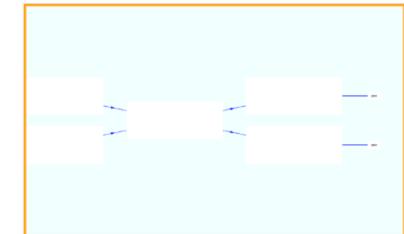
G08606CV



G81459MV

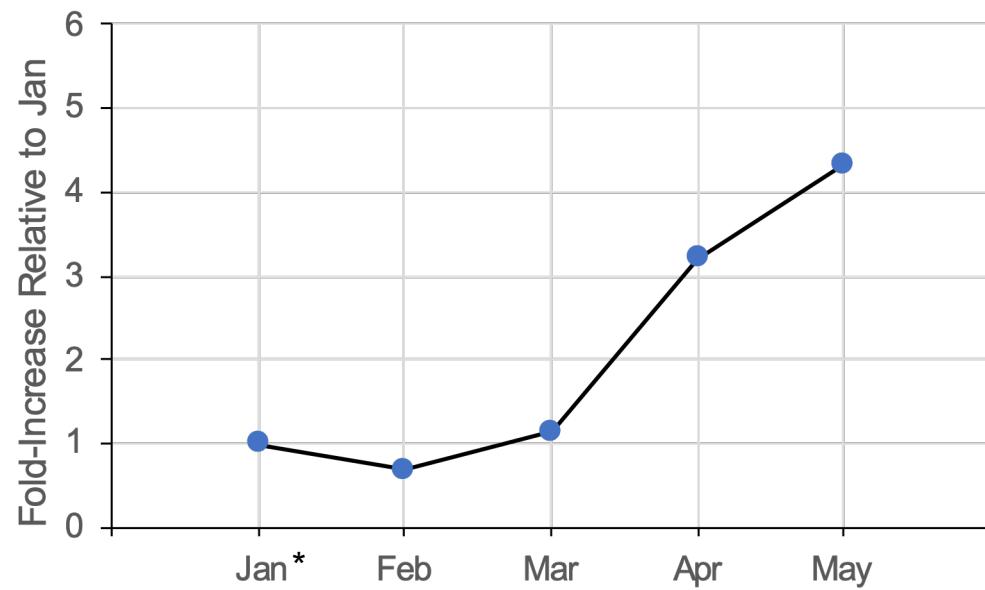
GNOME

GNOME - Glycan Naming and Subsumption Ontology

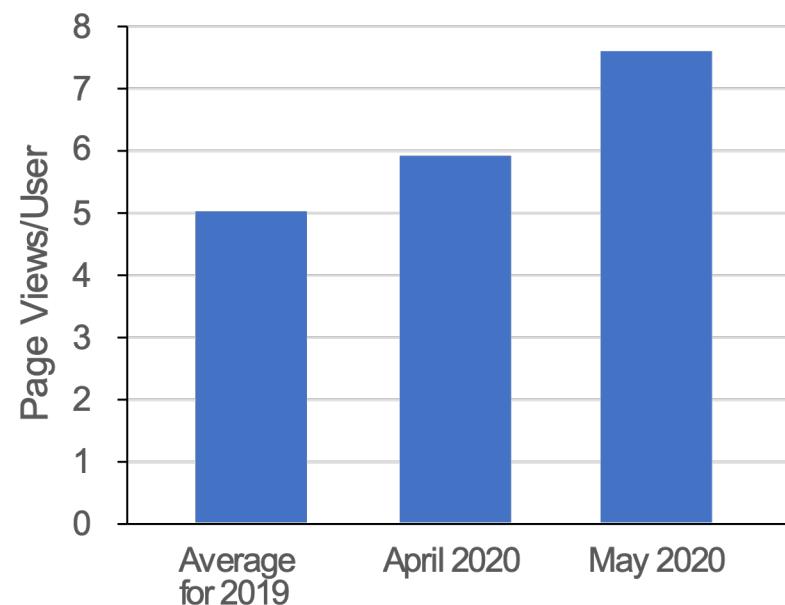


- **Improved mobile viewer**
- **Media page with slides, posters, etc.**
- **“How to Cite” page with past and new papers**
- **Expanded download options**
- **Expanded search options (PMID, multiple glycan ID namespaces, batch search for glycan and protein IDs, GO id, organism)**
- **v1.6, June 2020; Major migration of frontend from proprietary HTML to the REACT framework**
 - **Broad user and developer base (Facebook, Dropbox, Airbnb, Netflix, etc.)**
 - **Easier to add new components and reuse existing components in new contexts**

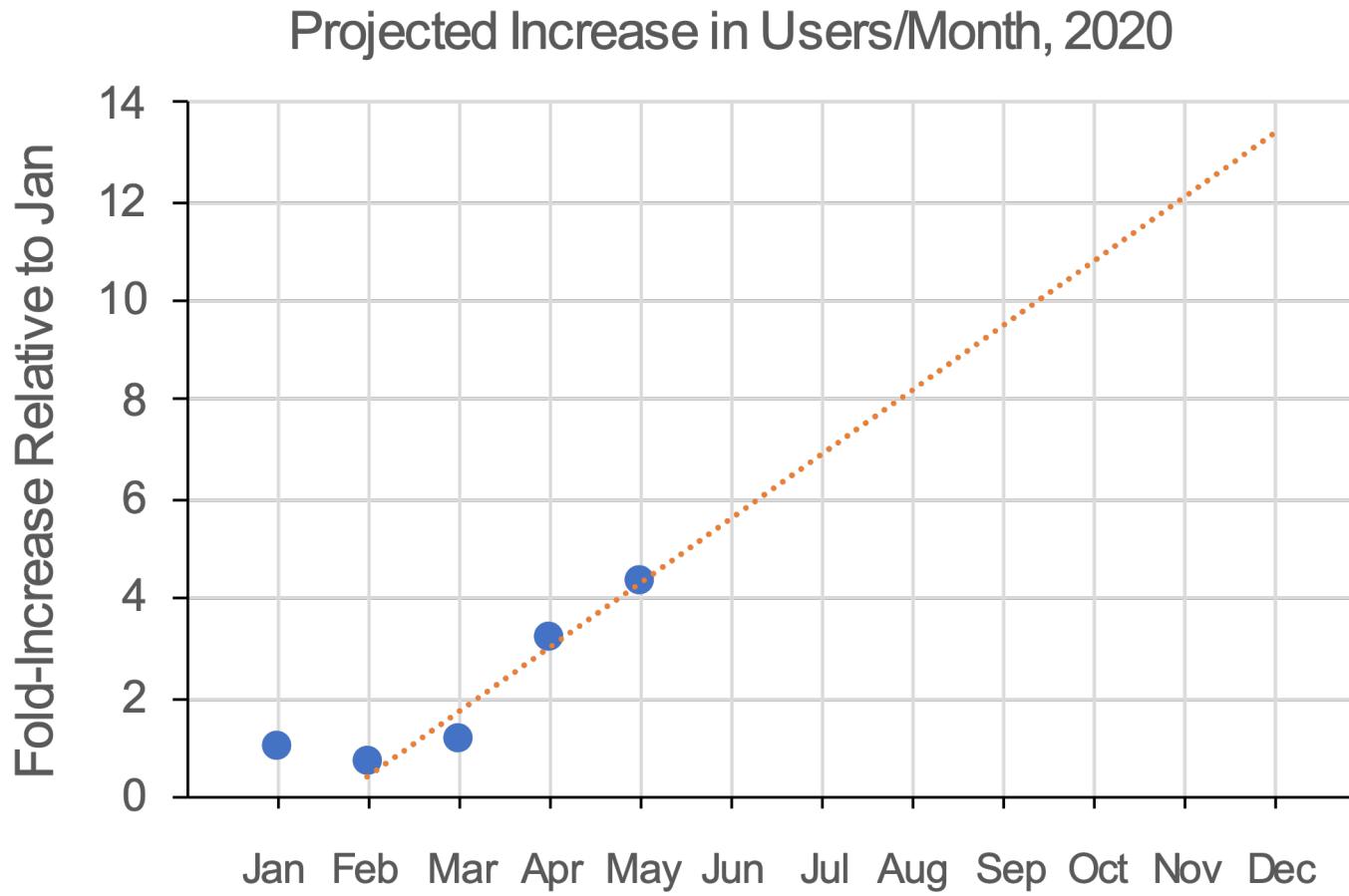
Increase in Users/Month, 2020



Increase in Page Views/User



*Note: Jan. 2020 user number higher than average of all months in 2019



GlyGen

Progress in Outreach

- **Demos (2019-2020)**

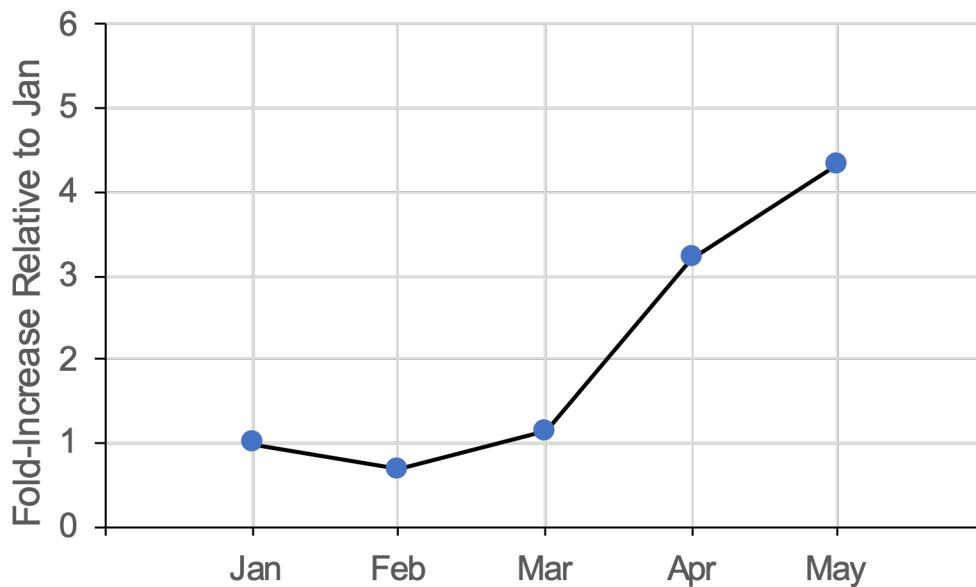
- GlyGen webinar at NIH-FDA Glycoscience Day, May 15, 2020
- GlyGen webinar for SIB glycobiology course May 11, 2020
- GlyGen Monthly meeting webinar on May 5th 2020
- Online demo given to PI and researchers at the University of Alabama, Birmingham March 2020
- GlyGen Demo at Demo at FDA White Oak Campus in March 2020
- Demo at George Mason University 26th Feb 2020
- GlyGen Demo at Exploring the Biology of O-GlcNAc an NIH Common Fund Glycoscience Program Symposium in March
- GlyGen Demo at Jeff Gildersleeve's Lab – January 6, 2020
- Glycoscience Commonfund Tools Subgroup Meeting - December 17, 2019
- GlyGen Demo at Rado Goldman's Lab (GU University) - October 10, 2019
- GlyGen Demo and assignment on GlyGen for GW students – September 10, 2019

- **Other posters (4) and talks (4) given, many cancelled (>6), including international training initiatives with GlySpace Alliance**

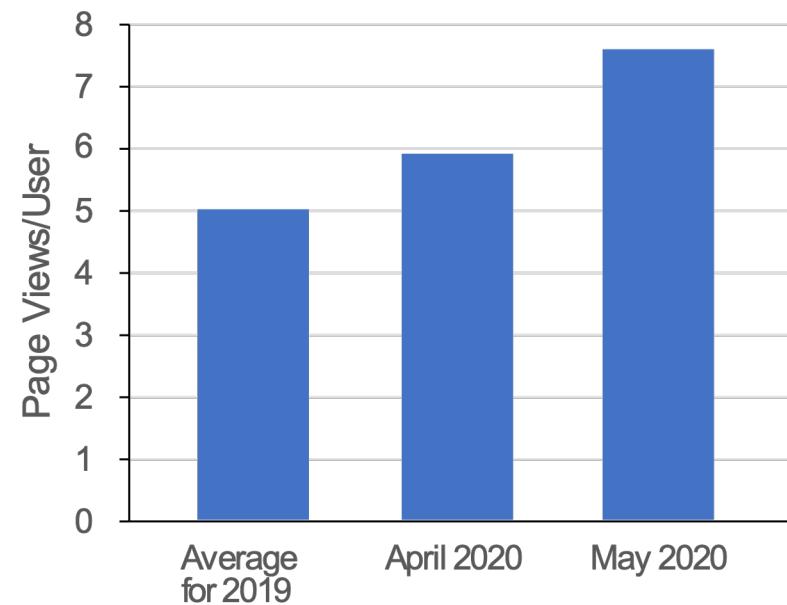


The value of reaching out and sharing the usefulness of GlyGen through demonstrations and webinars has become strikingly clear.

Increase in Users/Month, 2020



Increase in Page Views/User



General Announcement

ACCEPTED MANUSCRIPT

GlyGen: Computational and Informatics Resources for Glycoscience

William S York ✉, Raja Mazumder, Rene Ranzinger, Nathan Edwards, Robel Kahsay, Kiyoko F Aoki-Kinoshita, Matthew P Campbell, Richard D Cummings, Ten Feizi, Maria Martin ... Show more

Glycobiology, cwz080, <https://doi.org/10.1093/glycob/cwz080>

Published: 16 October 2019 Article history ▾

GLYCOBIOLOGY

Data manuscript

ACCEPTED MANUSCRIPT

GlyGen data model and processing workflow ⓘ

Robel Kahsay ✉, Jeet Vora, Rahi Navelkar, Reza Mousavi, Brian Fochtman, Xavier Holmes, Nagarajan Pattabiraman, Rene Ranzinger, Rupali Mahadik, Tatiana Williamson ... Show more

Bioinformatics, btaa238, <https://doi.org/10.1093/bioinformatics/btaa238>

Published: 23 April 2020 Article history ▾

Bioinformatics



PDF



Split View



Cite



Permissions



Share ▾

GlySpace Announcement

CORRECTED PROOF

The GlySpace Alliance: toward a collaborative global glycoinformatics community

Kiyoko F Aoki-Kinoshita ✉, Frederique Lisacek, Raja Mazumder, William S York, Nicolle H Packer

Glycobiology, cwz078, <https://doi.org/10.1093/glycob/cwz078>

Published: 01 October 2019 Article history ▾



GlyGen



GLYCOBIOLOGY

GlyGen Other Manuscripts (In Progress)

- Integration of GlyGen Glycans in ChEBI (lead – Gareth Owen)
- Curation of UniCarbKB-GlyGen data (lead – Matthew Campbell)
- GNOme (lead – Nathan Edwards)

GlyGen



Year 4 Goals

Future plans, data integration, feature development:

- **New data types:** additional species, natural variants, phenotypes, phosphorylation, glycan binding protein interactions (array data), protein-protein interactions, protein-GAG interactions, more glycosylation sites through automated literature mining
- **New annotations:** PTM functional annotations, more glycosylation subtypes, glycan names and relations (subsumption), biosynthetic pathways mapping enzymes to glycan precursor and products
- **New features:** enhanced filtering and sorting options, highlight hot-topics in glycoscience on the portal home page, new routes for accessing data through function/biosynthesis/disease, extensive help system using mediawiki, intuitive displays of data statistics, improved mobile friendliness

Positioning GlyGen for the future

- **Enhance Data Accessibility:** Present entry routes and access nodes to GlyGen datasets that are intuitive to the broadest possible range of investigators
- **Expand Data Volume:** Work with collaborators, individual investigators, and literature mining efforts to increase the dataspace accessible through GlyGen
- **Integrate Glycoconjugate Functions:** Link glycan and glycoconjugate structural data to diverse datasets that can support the generation of functional hypotheses and provide a mechanism for accessing glycan data through biological function
- **Outreach:** Capture the attention of biomedical investigators who are just realizing that they need a resource to help them address emerging glycoscience questions in their research domain

GlyGen will be offering three concurrent demonstration and discussion sessions in the afternoon. Each session will be repeated so that you will have the opportunity to attend two sessions of your choice.

Date: Tuesday 16th June, 2020 (Today)

Time: 4:00 – 4:45p.m and 5:00 – 5:45p.m

Session A. Use case discussions and data deposition

Session B. GlyGen Features Demonstration

Session C. Glycan ontology (GNOME), substructure searches and motifs

Join: Session A. Use case discussions and data deposition

- ❖ If you have a use case that you would like to see on GlyGen
- ❖ If you have glycobiology data and would like to share it with community through GlyGen
- ❖ If you have a need that you would like GlyGen to implement in future developments

Session A Webex information: Please remain logged into main meeting

Meeting link: <https://nih.webex.com/nih/j.php?MTID=m57d088f3e813b5a72cc8d23ca014e0df>

Meeting number: 126 729 7995

Password: Glyco2020

Join: Session B. Use GlyGen Features Demonstration

- ❖ If you are unfamiliar with what GlyGen can do
- ❖ If you want to get an overview of GlyGen's functionalities and features
- ❖ If you want to learn about how GlyGen can help you achieve your research goals.

Session B Webex information:

Meeting link: <https://nih.webex.com/nih/j.php?MTID=mbe6e4946fd2609183db76d5a06f2eb71>

Meeting number: 126 518 3438

Password: GlyGenDemo

Join: Glycan ontology (GNOME), substructure searches and motifs:

- ❖ If you want to learn structure-based strategies for finding
 - ❖ Glycan structures in GlyGen
 - ❖ “Related glycans...” in GlyGen
 - ❖ Glycan motifs, Glycoenzymes
 - ❖ GNOme Structure Browser

- ❖ If you want to learn strategies for finding glycan
 - ❖ Compositions in GlyGen
 - ❖ Mass-based advanced Glycan search
 - ❖ GNOme Synonyms in Ontobee
 - ❖ GNOme Composition Browser

Session C Webex information:

Meeting link:

<https://nih.webex.com/nih/j.php?MTID=ma12ac3336d60b685e107c313cd9b6ecb>

Meeting number: 126 022 6113

Password: VUpxDdvY599



Harvard University
Richard Cummings
Akul Mehta

The Jackson Laboratory
Judith Blake

Soka University
Kiyoko Aoki-Kinoshita

The Griffith University
Matthew Campbell
Jodie Abrahams

University of Delaware
Vijay Shankar
Peng Su

Imperial College London
Ten Feizi
Yukie Akune

Macquarie University
Nicki Packer

University of Georgia

Michael Tiemeyer
René Ranzinger
William York
Michael Pierce
Robert Woods
Rupali Mahadik
Tatiana Williamson
Sena Arpinar
Sanath Bhatt
Sujeet Kulkarni
Sandeep Nakarakkommula
Gaurav Agarwal
Vinamra Jain

EMBL-EBI

Maria Martin
Preethi Vasudev
Leyla Jael Garcia Castro

NIH-NCBI-NCI

Kim Pruitt
Evan Bolton
Jefferey Gildersleeve
Paul Thiessen
Terrence Murphy

The George Washington University

Raja Mazumder
Robel Kahsay
Jeet Vora
Rahi Navelkar
Nagarajan
Pattabiraman
Gareth Owen
Xiying Ding
Reza Mousavi
Xavier Holmes
Brian Fochtman
Tyler Stewart

Georgetown University

Nathan Edwards
Radoslav Goldman
Darren Natale
Karen Ross
Wenjin Zhang

GlyGen



Extra Slides

FAIRsharing (<https://fairsharing.org>), an informative and educational resource that describes and interlinks community-driven standards, databases, repositories and data policies to maximize the visibility and adoption of these resources



FAIRsharing Databases: A catalogue of databases, described according to the [BioDBcore guidelines](#), along with the standards used within them; partly compiled with the support of Oxford University Press ([NAR Database Issue](#) and [DATABASE Journal](#)).

databases > bsg-d001375

R GlyGen: Computational and Informatics Resources for Glycoscience

Abbreviation:GlyGen

General Information

GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data. GlyGen retrieves information from multiple international data sources and integrates and harmonizes this data. This web portal allows exploring this data and performing unique searches that cannot be executed in any of the integrated databases alone.

Homepage <https://www.glygen.org/>

Countries that developed this resource [United States](#)

Created in 2018

Taxonomic range

[Homo sapiens](#) [Mus musculus](#) [Rattus norvegicus](#)

Knowledge Domains

[Glycosylated Residue](#) [Glycosylation](#)

Subjects

[Glycomics](#) [Life Science](#)

User-defined Tags

[Glycobiology](#) [Glycoconjugate](#) [Glycoinformatics](#) [Glycoprotein](#) [Glycoscience](#)

Data License



Attribution 4.0 International (CC BY 4.0)

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially.
- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

Software/Source code License



The GNU General Public License v3.0

Free as in Freedom

- Freedom to use the software for any purpose,
- Freedom to change the software to suit your needs,
- Freedom to share the software with your friends and neighbors, and
- Freedom to share the changes you make.

GitHub Repository for Source Code

<https://github.com/glygener>



GlyGen is 5★ Open Data compliant

A rating system for open data proposed by **Tim Berners-Lee**, founder of the **World Wide Web (WWW)**



To score the maximum five stars, data must - <https://5stardata.info/en/>

Stars	Requirement	Fulfilment
1★	Available on the web (whatever format) <i>but with an open license, to be Open Data</i>	glygen.org (CC4.0, GPLGNU v3.0)
2★	Available as machine-readable structured data (e.g. excel instead of image scan of a table)	Datasets in csv format (data.glygen.org)
3★	as (2) plus non-proprietary format (e.g. CSV instead of excel)	All datasets in csv format
4★	All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff	Data available in RDF format via SPARQL query through GlyGen Triplestore (sparql.glygen.org/)
5★	All the above, plus: Link your data to other people's data to provide context	GlyGen links UniProtRDF, GlyCoRDF, PRO ontology and GlyCoO ontology

- 1★ glygen.org uses open public licenses
- 2★ All dataset are in machine readable format
- 3★ All the data in non-proprietary format such as csv, fasta
- 4★ GlyGen data can also be queried by using GlyGen's SPARQL endpoint (<https://sparql.glygen.org/>)
- 5★ GlyGen data is linked to other RDFs and ontologies

Twitter



GlyGen
@gly_gen

GlyGen is computational and informatics resources for glycoscience research. It allows for exploring carbohydrate related data and performing unique searches.

glygen.org

Joined February 2018

8 Photos and videos

Tweets 42 Following 7 Followers 104 Likes 1

Tweets **Tweets & replies** **Media**

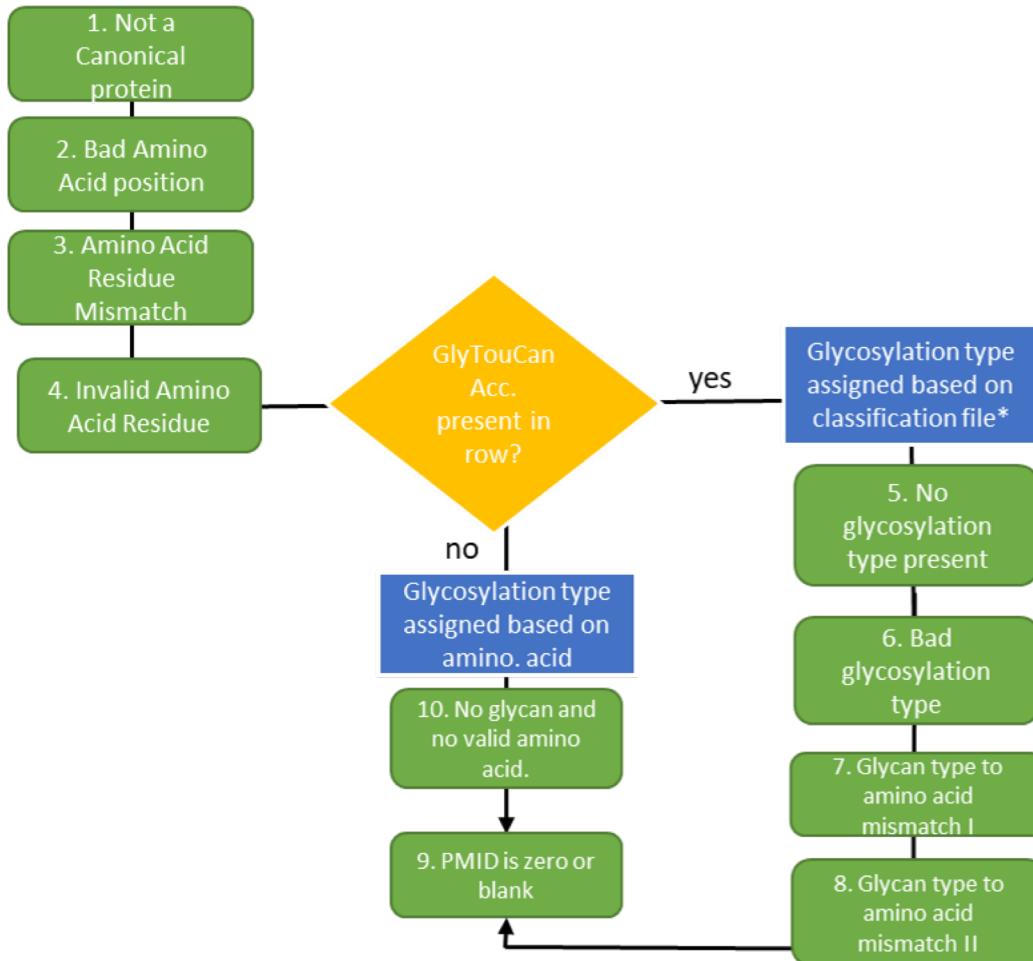
GlyGen @gly_gen · Dec 10
GlyGen has been added to the Common Fund Glycoscience Homepage (commonfund.nih.gov/glycoscience)



1 2 10

Monthly Talk

- ❖ GPTwiki: Glycopeptide Transitions for Targeted LC-MS Quantitation of Site-Specific Protein Glycoforms by Nathan Edwards (December 2019)
- ❖ GlycoDomainViewer - Bioinformatic tools for glycoproteomics by Hiren Joshi (August 2019)
- ❖ NIDCR's Human Salivary Proteome (HSP) Wiki database by Stefan Ruhl and William Lau (May 2019)
- ❖ NCBI's Conserved Domain Annotation Services by Aron Marchler-Bauer (Feb 2019)
- ❖ Mapping, analyzing, and visualizing PTMs in Jupyter Notebooks by Peter Rose (Jan 2019)
- ❖ Using canonical trees to assign glycoenzymes to glycan structures by Dr. William York and Dr. Alison Nairn (Oct 2018)
- ❖ Reactome - Pathway Context and Visualisation for Omics Data by Henning Hermjakob (Sept 2018)
- ❖ Current status and future plan of carbohydrate structures at the PDB by Jasmine Young (June 2018)
- ❖ Introduction to the Molecular Glycophenotype Ontology (MGPO) by Jean-Phillipe Gourdine (May 2018)

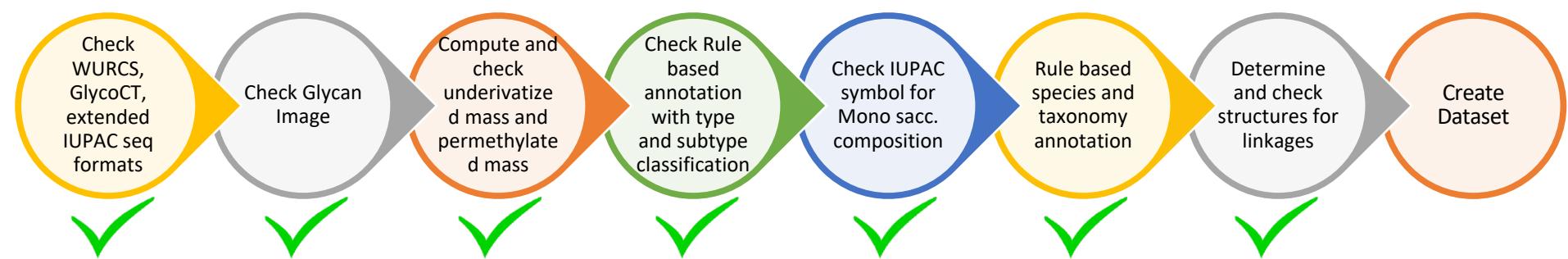


Check	Check_name in log file.
1	no_canon
2	bad_pos
3	aa_mismatch
4	invalid_aa
5	glycan_without_glytype
6	bad_glytype
7	n_glycan_aa_mismatch
8	o_glycan_aa_mismatch
9	no_pmid
10	no_glycan_invalid_aa

*classification file is generated by Nathan Edwards for GlyTouCan Ac.

GlyGen

Glycan Data Checks and annotations



- All protein accessions should have fasta sequence ✓
- All sequences should have MD5checksum ✓
- All accessions have a status: Reviewed or Unreviewed ✓
- All accessions are categorized: canonical or non-canonical ✓
- All annotations should have evidence and eco ids wherever applicable ✓
- All annotations assigned to amino acid should have a start and end range position/s ✓
- All annotations referring to position and amino acids should match the protein sequence ✓

GNOME Composition Browser: Hex X +

https://gnome.glyomics.org/GNOME.browser.html?HexNAc=4&Hex=5&dHex=1&NeuAc=2

G25570MS G53916LT G36220ML G08835ES G58006TV

G99129XB G92757GR G97666VC G74708PX G22561YO

G49904RA G93196BP

GNOME - Glycan Naming and Subsumption Ontology

The screenshot shows the GNOME Structure Browser interface. On the left, a vertical stack of colored triangles indicates the composition of each glycan structure. The colors represent different monosaccharides: blue for GlcNAc, yellow for Glc, green for Gal, red for NeuAc, purple for Fuc, and grey for sialic acids. The numbers next to the triangles indicate the count of each type. The main area displays ten glycan structures, each with a unique identifier (e.g., G25570MS, G53916LT) and a corresponding chemical structure diagram. The structures consist of various linkage patterns of these monosaccharides.

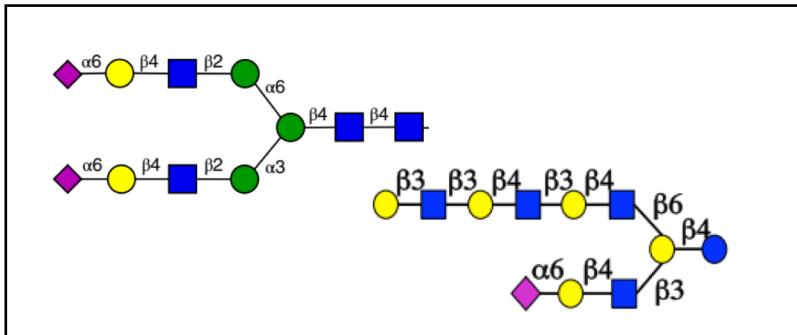
GNOme Composition Selector: Hex

<https://gnome.glyomics.org/GNOme.compositionselector.html?HexNAc=4&Hex=5&dHex=1&NeuAc=2>

The screenshot shows the GNOme Composition Selector interface. On the left, a vertical stack of colored boxes (blue, yellow, green, white, red, purple) with numbers (0, 0, 0, 4, 0, 0, 5, 0, 1, 2, 0, 0, s, 0) and symbols (triangles, circles, squares, diamonds) indicates the composition of a glycan. The main area displays four glycan structures (G42039FI, G45395BF, G64824NL, G17664QO, G44874SK) each composed of various shapes (circles, squares, triangles, diamonds) with question marks, representing unknown or variable counts. A legend at the bottom defines the symbols: blue triangle (top), blue square (middle), blue circle (bottom), yellow square (top), yellow circle (middle), yellow triangle (bottom), green square (top), green circle (middle), green triangle (bottom), grey circle (top), grey square (middle), grey triangle (bottom), red square (top), red circle (middle), red triangle (bottom), purple diamond (top), purple square (middle), purple circle (bottom), and a small blue square at the bottom center.

THE BOONS GROUP

Mapping or registering
synthesizing glycan
structures in
GlyTouCan database



GlyGen

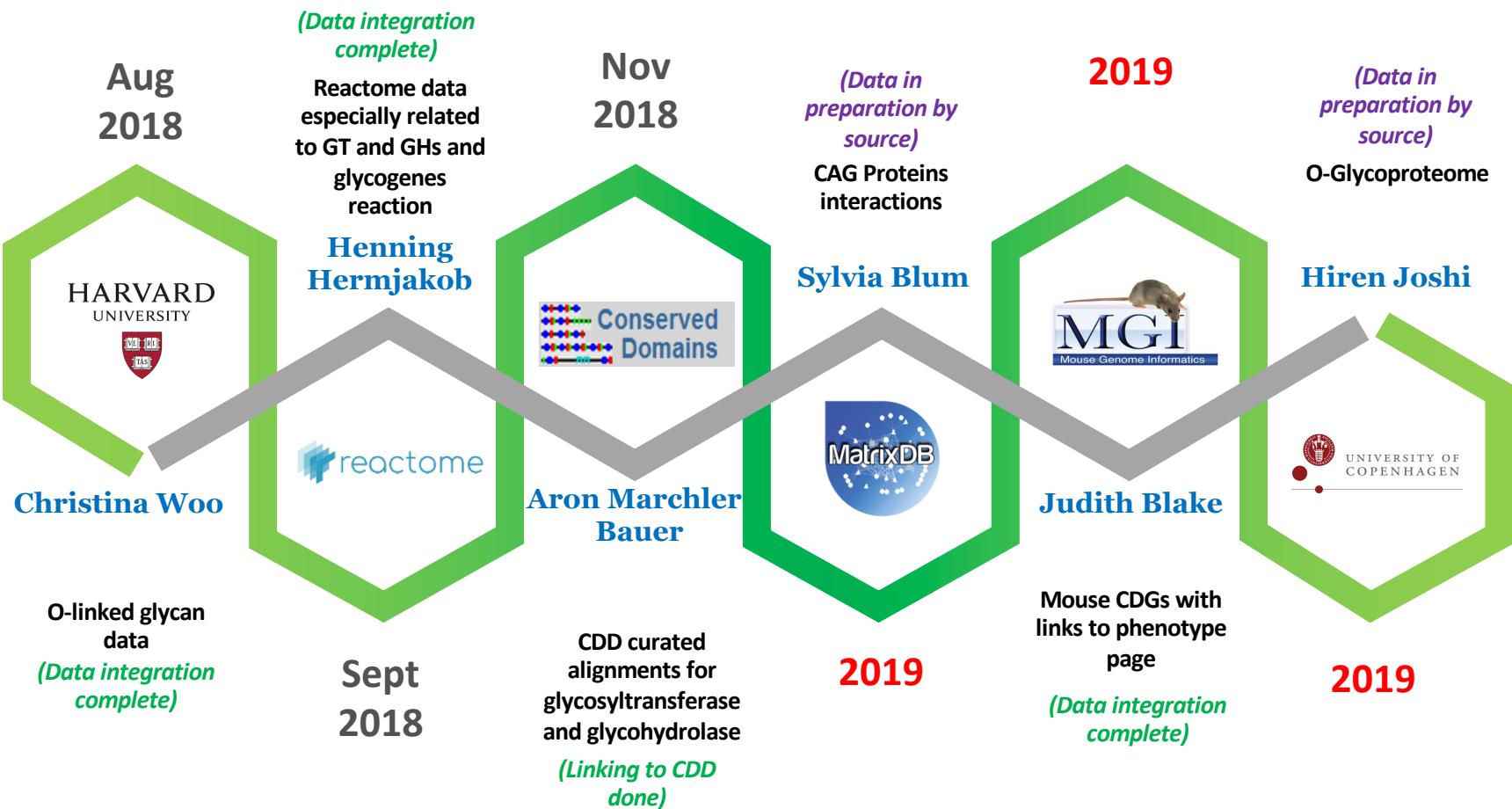
Submissions

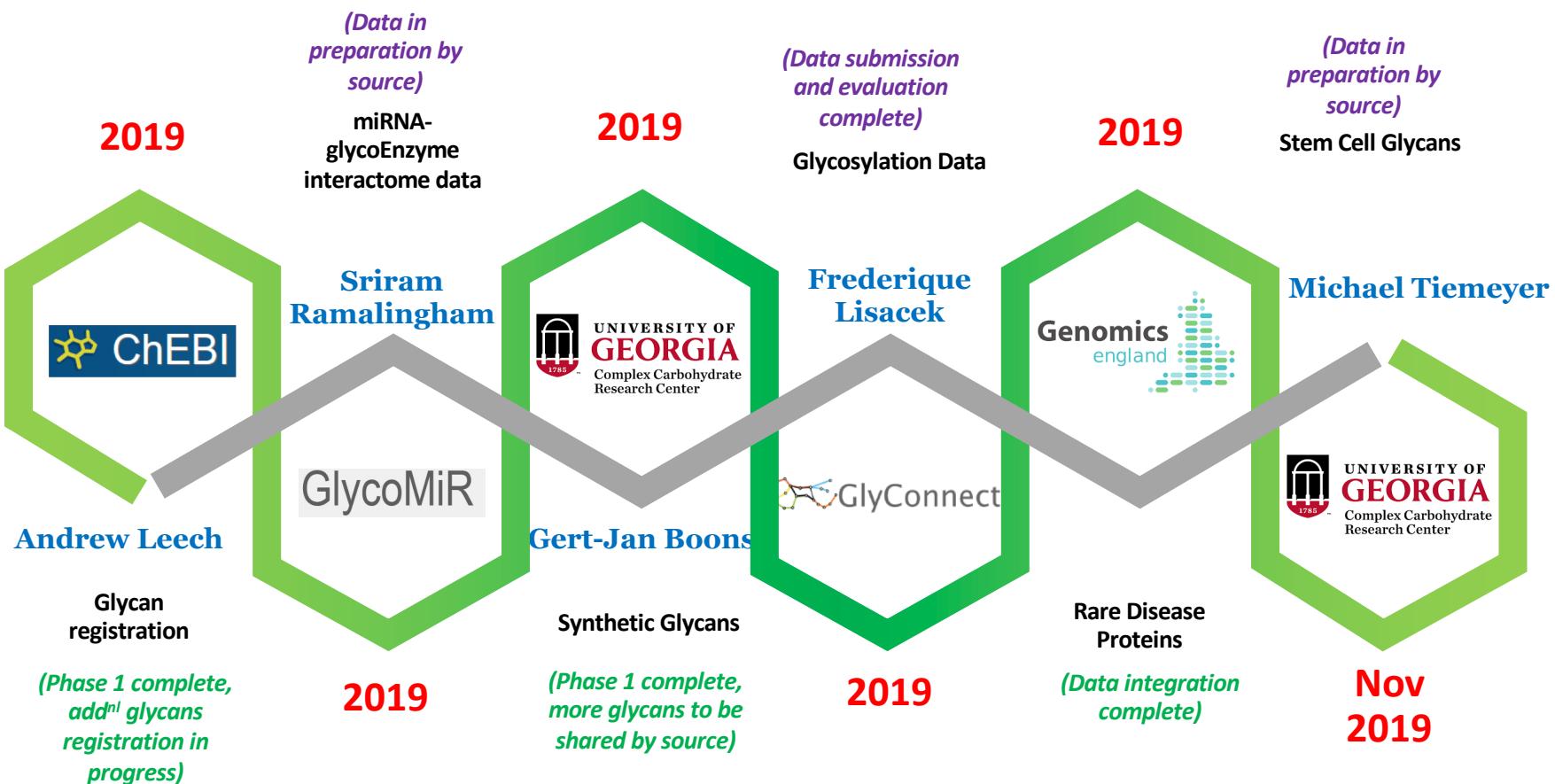
Registered ID	Accession Number	Sequence	Registration Date	Validation
9487abc519...	G350330R	WURCS=2,0/5,14,13/[a2122h-1x_5j[a2112h-1b_1-5j][a2122h-1b_1-5,2^*NCC3=O][Ad21122h-2a_2,6,5^*NC3=O]m1-2122h-1x_5j[1,3,3,2,3,3,2,5-3,2-3,2/a-b1,b3-c1,b5-f1,c4-d1,d6-e2,f4-g1,g3-h1,g6-k1,h4-i1,j1,k4-i1,l3-m1,m2-n1]	2019/08/13	Validation passed
sehccosbd4...	G79420BT	WURCS=2,0/4,11,10/[a2122h-1x_5j[a2112h-1b_1-5j][a2122h-1b_1-5,2^*NCC3=O]j1-2-3,2-a3-2,3-2-3-2/a-b1,b3-c1,b6-f1,c4-d1,d6-e2,f4-g1,g3-h1,h4-i1,j1,k3-k1	2019/08/13	Validation passed
6525ca1455...	G19289JN	WURCS=2,0/7,20,19/[a2122h-1x_5,2^*NCC3=O][a2122h-1b_1-5,2^*NCC3=O][a21122h-1b_1-5j[a1122h-1x_5j][a21122h-2a_2,6,5^*NC3=O]j1-2-3,2-a3-2,3-2-3-2/a-b1,b3-c1,b6-f1,c4-d1,d6-e2,f4-g1,g3-h1,h4-i1,j1,k3-k1	2019/08/13	Validation passed

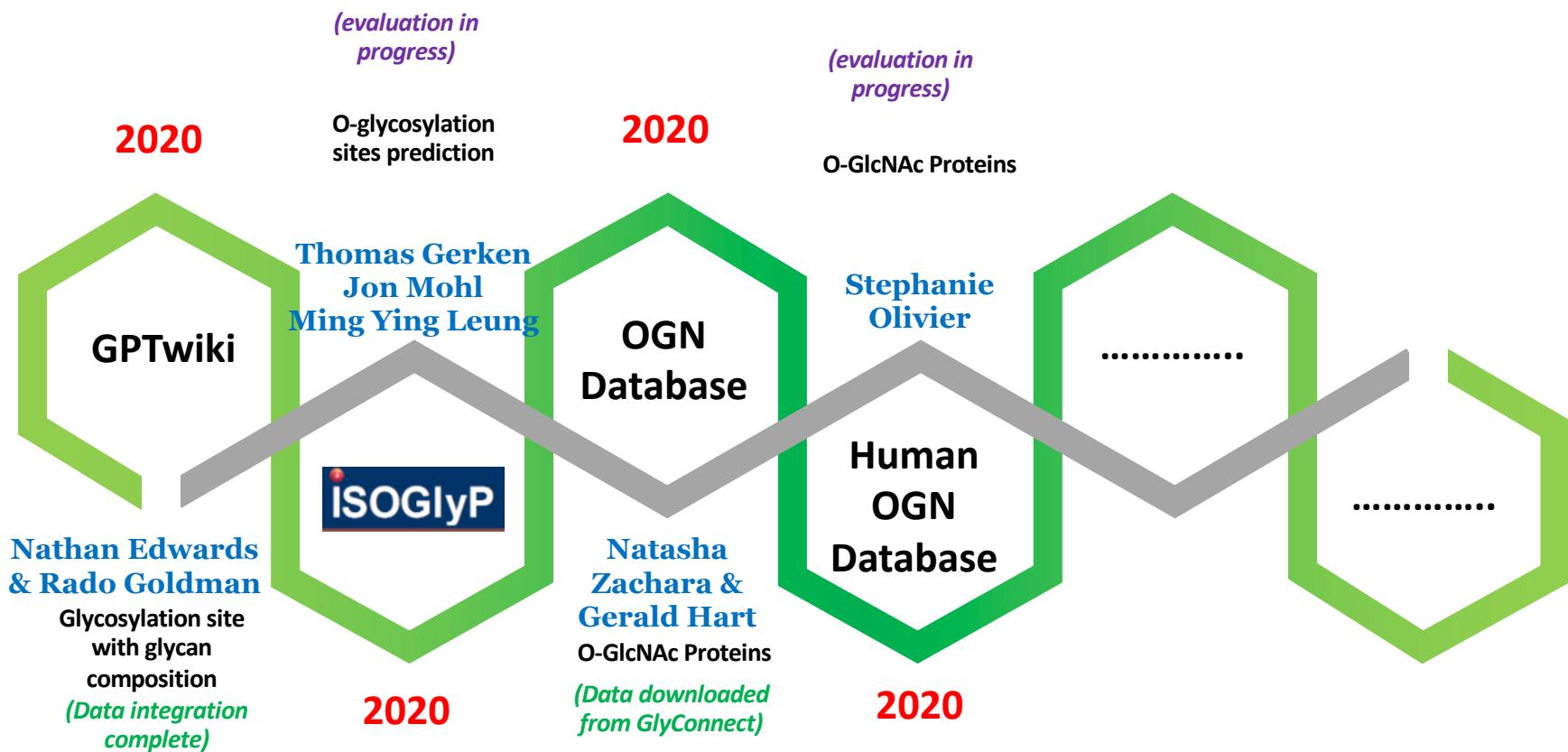
- Build a training team (post-doctoral level or other) that can provide remote support and on-site interactions with interested parties (Tiemeyer and Ranzinger)
- Accept invitation from Pamela Marino to help organize an NIH Intramural Training opportunity in which we present GlyGen and work with NIH Investigators to incorporate glyco-relevant datasets (Tiemeyer and Mazumder)
- “Meet a glycobiologist” roundtable at ASBMB this Spring, in conjunction with Tools Workshop at which GlyGen is presenting (Tiemeyer)
- Identify additional professional societies (glyco-knowledgeable and glyco-naïve) that are good targets for outreach: Society for Neuroscience, Mucin and Mucus meetings, ATS, CDG and other rare disease research meetings, ASCB, various GRCs, others TBD (Tiemeyer)
- Develop small user groups with various expertises, outside of the bioinformatics community, that provide regular feedback on features, failures, and unmet needs (Tiemeyer)

(Some In progress, more to follow)



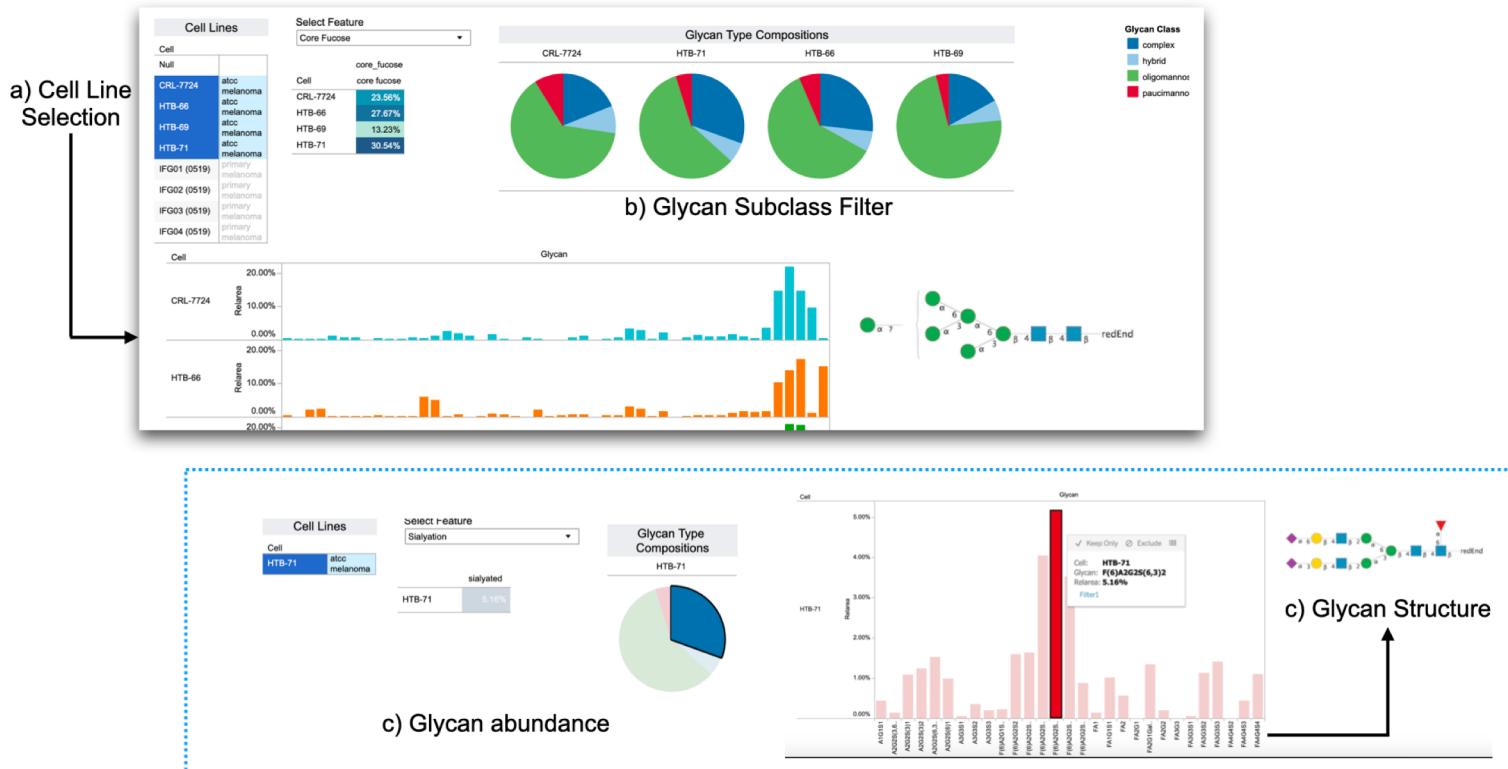






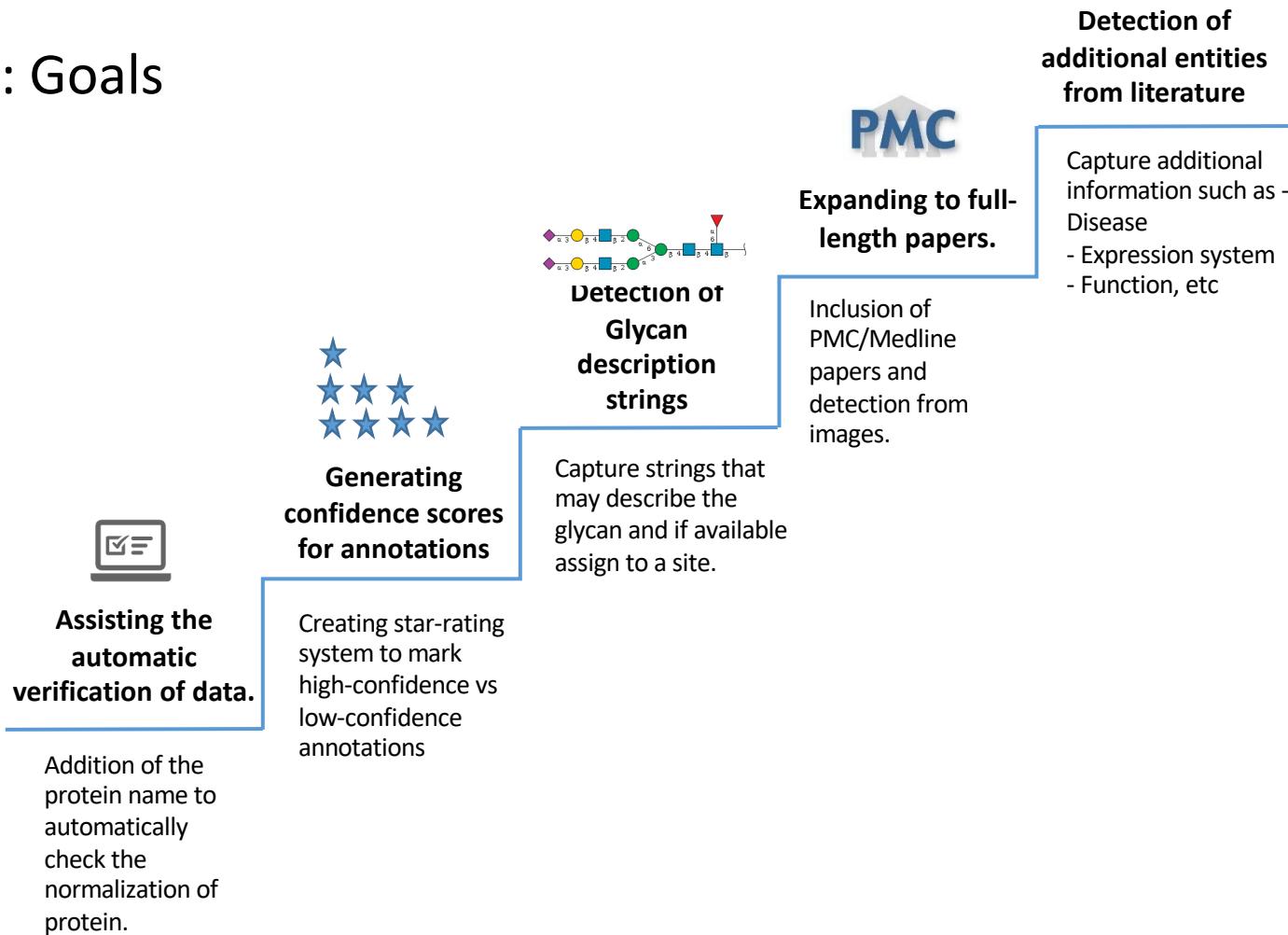
Modifications of specific sites in PRO:

- Before: came exclusively from PSI-MOD ontology
 - Modifier and amino acid residue form a single entity
 - Example, serine + phosphate => phosphoserine
 - Asn-Gly-(pSer)
 - The modified amino acid is part of the chain
- Issue: No similar treatment for glycans (thankfully!)
 - Modifier and amino acid residue are separate entities
 - Example, asparagine + glycan = asparagine~glycan
 - (Asn~glycan)-Gly-Ser
 - Modified amino acid is the 'base' amino acid attached to the modifier
- After: glycan modifications come from GlyTouCan via GNOme
 - new relation to indicate linkage between amino acid and glycan

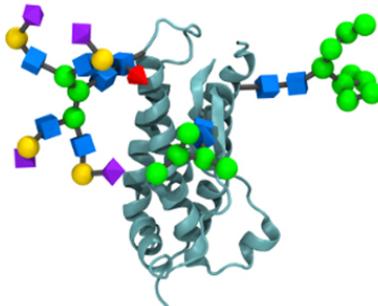


Data visualization tools for users to query and compare curated GlyGen data collections. Current releases designed for cancer cell lines and SARS-CoV-2

Year 2: Goals



Press releases



GlyGen allows you to search for glycoproteins based on proteins or glycan structures. Credit: GlyGen

GlyGen: Open data for glycoscience

4 Nov 2019 - 10:49

Summary

- GlyGen is a new informatics portal for glycoscience
- Glycans are carbohydrates that attach themselves to other molecules
- Understanding how glycans work could offer useful insights into human health and disease

4 November, 2019, Cambridge – Of the four building blocks of life – proteins, lipids, nucleic acids and glycans (carbohydrates) – the latter have received the least attention from researchers. This may be about to change as a new project, [GlyGen](#), launches to help researchers answer glycan-related questions using big data.

GlyGen is the first project that integrates and harmonises glycan data from multiple international data sources, making it much faster and simpler for researchers to explore the data and perform unique searches. The portal currently contains glycan data for three species: human, mouse and rat. More organisms will be added in the near future.

Newsletter

Glycomics, the systematic study of all glycan structures is central to understanding the macromolecular space within a given cell type or overall organism. In recent years remarkable innovation in the areas of ion mobility and high resolution mass spectrometry, ultra high pressure liquid chromatography and induced hyperpolarization methods for NMR have compounded the availability of diverse glycan structural information. In addition, high throughput analysis of glycan-protein binding interactions by way of specialized microarrays has allowed for rapid characterization of the estimated 10 to 100 million unique glycoproteins in the human body.

As the glycome is known to have a significant effect on the function of complex biological processes as well as many diseases and disorders including cancer, glycobiology is having an increasing impact on the missions of the NIH and FDA. In addressing the apparent lack of adequate glycoinformatic infrastructure, we present the GlyGen knowledgebase that integrates multidisciplinary data and knowledge that contribute to the conceptual basis for glycobiology. This includes knowledge about the molecular, biophysical and functional properties of glycans, genes, proteins and lipids organized in pathways and ontologies as well as a rapidly growing body of biological big data related to mutation and expression.

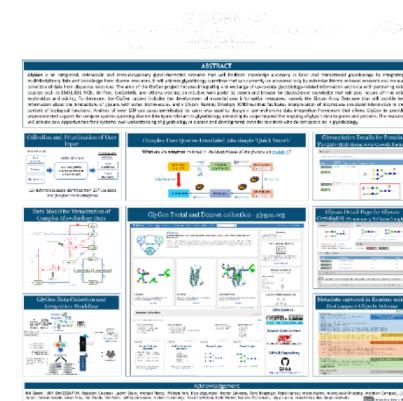
New features in the latest version of GlyGen include improvements to search capabilities as well as additional glycan-centric datasets. Specifically, updates to the API in this version facilitate search by monosaccharide composition, as well as batch retrieval linked to statistics across the glycan, protein, and glycoprotein domains for nuanced queries. Further developments include the inclusion of over 20,000 new glycans and Reactome data for exploration of pathways of human proteins. Moreover, the addition of specialized datasets from collaboration with UGA, Harvard, and UniprotKB cover niche areas relevant to researchers at the cutting edge of glycoscience such as congenital disorders of glycosylation and details relating to O-linked glycosylation.

[Paragraph About Future Plans]

Version 1.6 will have changes in portal design, redesigned glycosylation section, new sections in protein and glycan details page that will include ptm annotations, reactions and pathways, protein binary interaction details, addition of new usecases and datasets, addition of a new species, support for other download formats, a basket to add accessions and information for bulk download and much more.

- **Talks and Posters**

- CCRC UGA Seminar Talk 27th Feb, 2020
- Common Fund Tools Group Monthly Conference, Dec. 2019; talk and demo
- Society for Glycobiology 2019, Phoenix; 2 posters and ad hoc demo
- NIH Glycoscience Day Poster and demo - July, 2019
- Beilstein Symposium Talk - June 2019
- NIH Common Fund Glycoscience Talk and Demo - May 29-30, 2019
- NIH Open Science talk - Oct 31, 2018
- SFG New Orleans Talk - November 5-8, 2018
- Biocuration Talk Cambridge and Poster - April 7-10, 2018
- Warren Workshop Talk - August 2020





GlyGen

GlySpace Alliance

Goals of GlySpace Alliance

PMID: 31573039 DOI: 10.1093/glycob/cwz078

- ❖ Promote the **use of standards** for representing and exchanging glycobiology information;
- ❖ Develop protocols for **establishing and representing the provenance of glycobiology data** and acknowledgement of the individuals and groups that collect, curate, map and integrate this data;
- ❖ **Establish metrics regarding the reliability and accuracy of the data** based on existing metrics for quality control (e.g., evidence ontology);
- ❖ **Reach consensus for data filtering and/or annotation** when disparities are found among the data;
- ❖ **Provide globally distributed data storage redundancy to enhance data accessibility** and ensure its longevity

GlySpace Alliance Training Initiative



- Joined initiative of the 3 major glycomics database providers
 - ExPASy server (Swiss Institute of Bioinformatics)
 - GlyCosmos (Soka University)
 - GlyGen (UGA + GW)
- Two days guided handson training meetings
- Each GlySpace member will send tutors to the training meetings to teach attendees about the corresponding resource
- Three training opportunities planned in 2020 (**Will have to be postponed due to COVID-19**)
 - Washington DC, USA: in conjunction with the Annual Meeting
 - Grenoble, France: in conjunction with the Warren Workshop 2020
 - Tokyo, Japan: in conjunction with the annual Japanese Society of Carbohydrate Research meeting
- Additional meetings for 2021 will be planned once the first meetings have been successful

GlyGen